

h

UMANISTISKE

DATA



Språk
Kultur
Optiske
plater



**NAVS EDB-SENTER
FOR HUMANISTISK FORSKNING**
NORWEGIAN COMPUTING CENTRE
FOR THE HUMANITIES

3-87

SENTERETS RAPPORTSERIE

RAPPORTER UTGITT F.O.M. 1980

RAPPORT nr. 17. *Svein Lie: Automatisk syntaktisk analyse*. Del 1. Grammatikken. Desember 1980. ISBN 82-7283-014-0 Pris kr. 30.

RAPPORT nr. 18. *Datateknologi og humanistisk forskning*. Bidrag til en NAVF-utredning. Desember 1980. ISBN 82-7283-015-9 Pris kr. 30.

RAPPORT nr. 19. *Statistiske metoder på arkeologisk materiale*. Rapport fra et seminar på Bryggens museum, Bergen 24.-26. november 1980. Mars 1981. ISBN 82-7283-017-5 Pris kr. 35.

RAPPORT nr. 20. *EDB-prosjekter i humanistiske fag 1980*. Juni 1981. 2. opptrykk oktober 1981. ISBN 82-7283-018-3 Pris kr. 45.

RAPPORT nr. 21. *Rune Johansen: Bruk av EDB i teatervitenskapelig forskning*. Mai 1981. ISBN 82-7283-019-1 Pris kr. 35.

RAPPORT nr. 22. *Årsmelding 1980*. NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-020-5 Gratis.

RAPPORT nr. 23. *Stig Welinder: A program package for archaeological use*. 1981. ISBN 82-7283-021-3 Pris kr. 45.

RAPPORT nr. 24. *Rapport fra seminar om bruk av edb innen teater og teatervitenskap*. Januar 1982. ISBN 82-7283-026-4 Pris kr. 50.

RAPPORT nr. 25. *Ole Lauvskar: Diskriminantanalyse i SPSS*. Desember 1982. ISBN 82-7283-028-0 Pris kr. 55.

RAPPORT nr. 26. *Stig Welinder: Paleodemography*. Oslo 1982. ISBN 82-7283-030-2 Pris kr. 55.

RAPPORT nr. 27. *Årsmelding 1981*. NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-029-9 Gratis.

Forts. 3. omslagsside.

HUMANISTISKE DATA

3-87

***NAVF's EDB-SENTER
FOR HUMANISTISK FORSKNING***

***The Norwegian Computing Centre
for the Humanities***

HUMANISTISKE DATA

utgis av NAVFs edb-senter for humanistisk forskning.

REDAKSJON

Jostein H. Hauge, Anne Lindebjerg, Kristin Natvig (red.), Espen S. Ore.

ADRESSE

Harald Harfagresgt. 31. Boks 53 - Universitetet, 5027 Bergen. Tlf. 05-212954/55/56.

ABONNEMENT

Abonnementspriser f.o.m. 1988: kr. 50,- for enkeltpersoner, kr. 100,- for institusjoner (3 nr. pr. år). Gratis for abonnement utenlands.

Artikler, rapporter og meldinger mottas gjerne.

Medarbeidere fra Senteret i dette nummer:

Reidun Oanæs Andersen, Kristin Natvig, Eli Schilbred, Øystein Reigem.

Redaksjonen avsluttet 1. desember.

Forsidebildet: Det italienske firmaet SIDAC har utviklet optiske lagringssystemer for bl.a. kunstsamlinger, og er i gang med flere omfattende prosjekter. Se Ernesto Bartolozzi artikkel.

NAVFs EDB-SENTER FOR HUMANISTISK FORSKNING

ble opprettet av Norges allmennvitenskapelige forskningsråd i 1972. Senteret skal arbeide på nasjonal basis for bruk av edb i forskningsarbeidet i de humanistiske fagene.

Sentrale oppgaver er edb-tjenester (veiledning og betalte oppdrag) og program- og metodeutvikling. Senteret holder kurs, seminar og konferanser om bruk av edb i humanistiske fag. Foruten utgivelse av Humanistiske Data omfatter informasjonstjenestene en rapportserie, årsmelding og elektronisk informasjonsformidling.

Senteret er sekretariat og operativt edb-organ for International Computer Archive of Modern English (ICAME), og utgir tidsskriftet ICAME Journal. Senteret er ansvarlig for administrasjon og drift av Norsk tekstarkiv og har det administrative ansvar for NAVFs sentral for informasjon om forskningsprosjekter (SIF) og Fagttjenesten for informasjon om humanistiske forskningsprosjekter (SIF-H).

Humanistiske Data is published by The Norwegian Computing Centre for the Humanities. Editorial group: Jostein H. Hauge, Anne Lindebjerg, Kristin Natvig (ed.), Espen S. Ore.

The journal can be ordered free of charge from the address above. Contributions are welcome. On request the Centre can supply the addresses of contributors to the journal.

GRAFISK FORMGIVNING: Svart på Hvitt.

SATS: Gerhard Datasats A.S.

TRYKK: Bergen Trykk A.S.

INNHOOLD

Redaksjonelle data	4
--------------------------	---

ARTIKLER

BBC Advanced Interactive Video and the Domesday Discs. <i>Phyllis S. Gove</i>	6
The North West Educational Computing Project. <i>Ian Robertson</i> <i>og Mike Picciotto</i>	15
Recording the Italian Cultural Heritage. <i>Ernesto Bartolozzi</i>	22
A CD-ROM-Based Geographic Information System. <i>Erling</i> <i>Maartmann-Moe</i>	33
Framandspråkundervisning og data. <i>Bjørn Dommarsnes</i>	42
Dialektekspert. <i>Torbjørn Nordgård</i>	50
Aktuelle metoder for automatisk morfologisk analyse og mulige anvendelser av dem i fremtidens tekstbehandlingssystemer. <i>Tove Fjeldvig</i>	64
Humanistisk datalogi i Danmark. <i>Mogens Baumann Larsen</i>	82

RAPPORTER

Interaktiv video i fremmedspråksundervisningen. <i>Reidun Oanæs</i> <i>Andersen</i>	89
RD-BIB. <i>Eli Schilbred</i>	99
Multi-Lingual Scholar TM versjon 3.0. <i>Torrey Seland</i>	107
Humanioradagene 1987. <i>Kristin Natvig</i>	115

MELDINGER	117
------------------------	-----

SUMMARY	124
----------------------	-----

REDAKSJONELLE DATA

Som leserne vel har merket, har Humanistiske Data fått en betydelig «ansiktsløftning» det siste året. Vi i redaksjonen håper dere synes at bladet er blitt både mer tiltrekkende og mer lesevennlig. Dessuten har opplagsøkningen vært gledelig stor i 1987 — bladet trykkes nå i 1400 eksemplarer.

Humanistiske Data har i alle år vært gratis for enkeltpersoner. Dette har vært mulig p.g.a. en enkel design, begrenset sideantall og et forholdsvis lite opplag. Men forbedringen av bladet sammen med opplagsøkningen har medført økte produksjonsutgifter. For øvrig har trykkerprisene øket i været i år.

Vi ser oss derfor nødt til å innføre en abonnementsavgift på **kr 50,- pr. år** for enkeltpersoner f.o.m. 1988. Avgiften for institusjonsabonnementer stiger til kr. 100,-. Abonnenter i utlandet vil fortsatt motta bladet gratis.

Vi mener at 50 kr er en lav pris å betale for over 300 sider aktuelt fagstoff i året. Avgiften dekker da heller ikke på langt nær våre faktiske produksjonskostnader. Likevel vil økte abonnementsinntekter gjøre det mulig for oss ikke bare å holde bladet på det nåværende nivå, men å gjøre det enda bedre.

Alle abonnenter vil altså motta en regning for Humanistiske Data i 1988. Skulle det være noen lesere som ikke ønsker å fortsette å abonnere på bladet, ber vi dem om å gi beskjed snarest.

Så noen ord om dette nummeret av bladet. De fire første artiklene er skrevet av foredragsholdere på konferansen «Tekst Lyd Bilde» som Senteret arrangerte 14.-16. oktober. I HD 1-88 vil leserne finne sammendrag av de øvrige foredragene som ble holdt på konferansen, samt en generell rapport fra arrangementet. Det andre hovedtemaet denne gangen er edb og språk, og også denne gangen kan vi gi programvare-anmeldelser.

Presentasjoner av programvare er nettopp en type stoff som vi vet at leserne ønsker mer av. Større vekt på praktiske emner i det hele tatt etterlyses også. Vi i redaksjonen skal i framtida gjøre vårt beste for å imøtekomme disse ønskene. Vi satser bl.a. på å gjøre programvare-anmeldelser til faste innslag.

For å lage et tidsskrift som er bedre tilpasset lesernes behov, trenger vi imidlertid også hjelp fra *dere*. Kanskje *du* har testet et program

som også andre humanister kan ha nytte av (eller bør holde seg unna)? Eller du 1. arbeider med et prosjekt, 2. har vært på et kurs/seminar/konferanse, 3. har lest ei nyttig bok osv. osv. som andre vil ha interesse og glede av å lese om. Vi setter stor pris på bidrag av alle slag — fra korte rapporter til lengre artikler, og i nivå fra det populære til det vitenskapelige. Bidrag kan gjerne være av kritisk eller provoserende art — det er mange emner og temaer innenfor fagfeltet som blir eller bør bli debattert. Spaltene er selvsagt også åpne for korte meldinger, leserbrev m.m. Og har du ikke tid til å skrive noe selv — tips oss, så kan *vi* gå videre med stoffet.

Vi tar også svært gjerne imot både ros og ris fra leserne. Er det noe du savner i bladet, eller har du noe annet å klage over, så ring eller skriv noen ord.

Vi ser fram til et fruktbart samarbeid med både lesere og bidragsytere i 1988.

Kristin Natvig
redaktør

BBC ADVANCED INTERACTIVE VIDEO AND THE DOMESDAY DISCS

Phyllis S. Gove

The BBC Domesday Project, originally conceived as a celebration of the 900th anniversary of William the Conqueror's Domesday Book, has involved up to a million people in the creation of a contemporary electronic portrait of the United Kingdom. The two Domesday Advanced Interactive Videodiscs and a specially developed AIV system offer users over a million screens of information in still and moving pictures, maps, statistical data sets and text in one easily accessible database.

Much of the material in the form of photos and text for the Community Disc was collected by more than 15,000 schools and community groups in 1985, and consequently much publicity was given to this effort. The National Disc, a resource base created by collaborative effort on the part of universities, governments, commercial and voluntary agencies and co-ordinated by the BBC, has received far less attention. Both discs form a substantial national archive with potential applications for a wide range of users in education, commerce, industry, government and information services.

We examine the development of the Domesday Project, the contents and capabilities of the Community and National Discs, the AIV technology and additional BBC resources for AIV.

THE DOMESDAY IDEA

Peter Armstrong, an executive producer for BBC Television and a microcomputer enthusiast, conceived the idea of the Domesday Project during the BBC's planning for a television series to mark the 900 years of William's Domesday Book. In order to create the necessary hardware and software for a modern Domesday Book in two years' time, Peter enlisted support from BBC Enterprises, Philips Electronics, Acorn Computers, and the Department of Trade and Industry; formed an Editorial Board from the BBC, university and government depart-

ment advisers; and drafted a Domesday Project team from within the BBC and sub-contractors in British universities and government departments. By December 1984 invitations to participate in a major data collection exercise were sent out to the 33,000 UK schools and Domesday was well under way.

THE DISCS

The accompanying chart in figure 1 presents an overview of the two videodiscs. The Community Disc is place-orientated and contains Ordnance Survey (OS) maps at various levels of scale with photographs and text relating to each level. The National Disc, however, is essentially topic-orientated with a wealth of statistical data sets, photographs, film, previously published text articles and especially commissioned essays relating to life in Britain in the 1980s.

THE COMMUNITY DISC

The table in Figure 2 shows the various levels of OS maps at which text and photographs are stored on the Community Disc. At levels one and two over 500 satellite images and 900 aerial photographs provide a pictorial representation of the UK; and the text at level two has been written by local area experts, members of the Geographical Association, or supplied by the Automobile Association.

Level three contains the OS 1:50 000 scale maps, known familiarly as the old one inch to one mile maps, and reports in photographs and text from more than 15,000 schools and community groups who took part as volunteers in the Domesday data collection process during 1985.

At level four approximately 70 city centres in the UK are covered by OS 1:10 000 maps. And finally at level five there are a limited number of floor plans of special feature sites, accompanied by photographs and text submitted by schools and community groups.

Each area of the country is covered in text and photographs in at least one level, namely level two, and potentially on other levels. The coverage of country at level three, by the schools and community groups, corresponds roughly to the 1981 census population distribution map with text and photographs for more than 9,000 of the possible 23,000 4 x 3 km blocks. In total the Community disc contains over 150,000 pages of text, 20,000 photographs and 24,000 photographs of the Ordnance Survey maps.

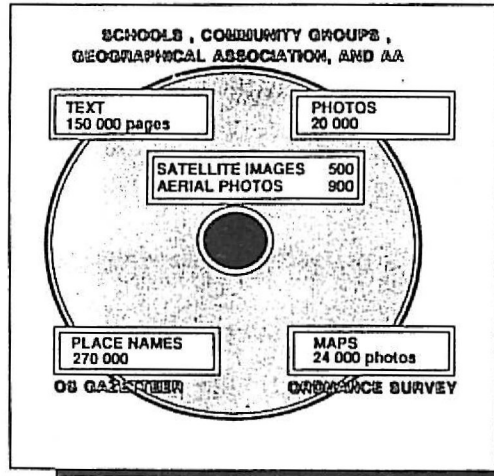
BBC ADVANCED INTERACTIVE VIDEO

THE DOMESDAY PROJECT

COMMUNITY DISC

enter through MAP/FIND

PLACES

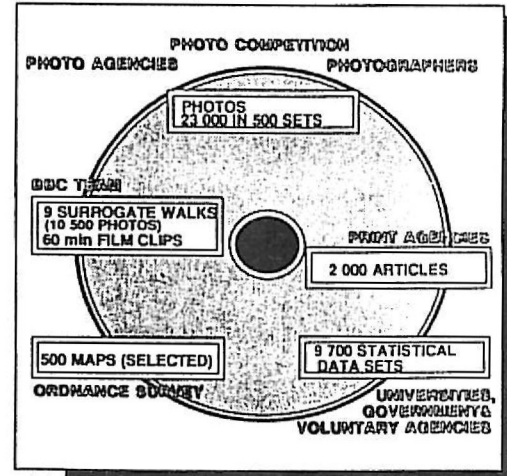


HARDWARE AND SOFTWARE DEVELOPMENT: 1985-1986

NATIONAL DISC

enter through GALLERY/CONTENTS/FIND

TOPICS



LV ROM = LASER VISION READ ONLY MEMORY

LV: Photos, maps and film are stored as VIDEO PICTURES on the disc
ROM: Text, captions, statistical data sets, and retrieval and indexing software

Fig. 1.

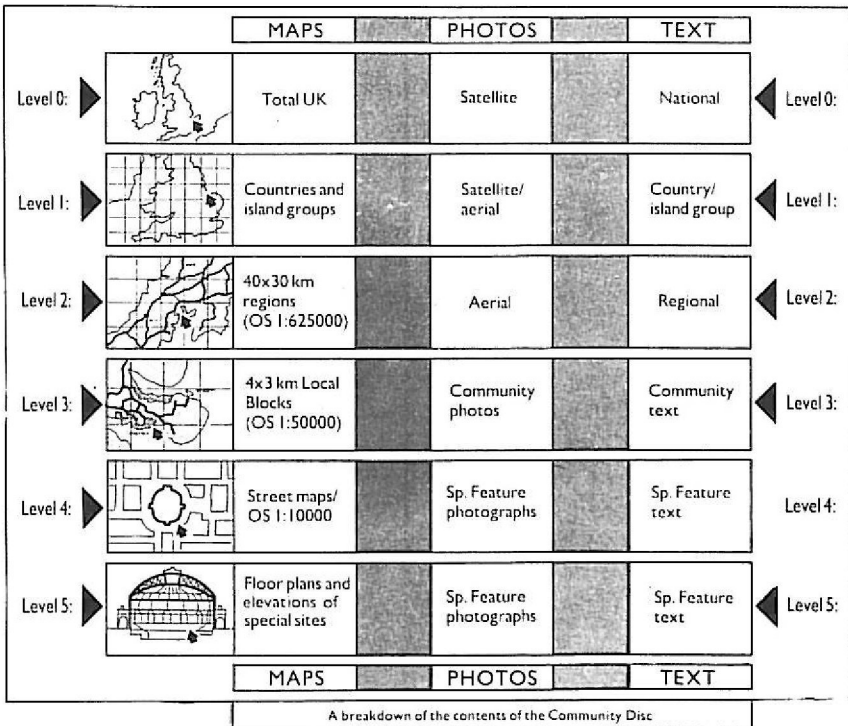


Fig. 2.

There are various means of accessing information on the Community Disc. A user may type in any one of over 7,000 subject keywords to the text and photographs; any one of over 270,000 place names provided by the OS Gazetteer or a grid reference. In addition by using a tracker ball control one can point to a location on a map and select the desired destination. Information choices can be easily selected from menu bars on the screen by the tracker ball control.

THE NATIONAL DISC

As well as providing a picture of life in Britain from the viewpoint of the local communities, the Domesday Project has also made information which is theoretically in the public domain more easily available on the National Disc. Referring again to the chart in figure 1 we note that the National Disc contents represent a major archive of over

23,000 photos in over 500 picture sets; nine surrogate walks that are made up of over 10,500 photos; 60 minutes of film with sound from BBC news and current affairs; 2,000 text extracts from previously published government reports, newspapers and journals; 45 specially commissioned essays which provide an overview of national subjects such as the environment or the armed forces; and 9,700 statistical data sets.

The visual information on the National Disc is a massive 'family album' of British life in the 1980s. The 500 picture sets were gathered from national photo agencies, top British professional photographers and a specially held national photographic competition. The nine 'surrogate walks' are a relatively new feature in interactive video, enabling a user to walk around an environment. The picture database for the walks comprises 10,500 photographs taken in a variety of representative houses and landscapes and urban areas. As a third pictorial resource the National Disc contains 60 minutes of moving film and sound sequences of the major news and events for each year of the 1980s so far.

The principal Government and institutional statistical data sources have made available to the BBC their main databases for publication on the National Disc. The main subcontractors, in alphabetical order, for this task have been the Department of Geography, Birkbeck College, the University of London; the Centre for Urban and Regional Development Studies and the Department of Geography, Newcastle University; the Economic and Social Research Council Data Archive, Essex University; and the Institute of Terrestrial Ecology, Bangor.

Almost all the statistical databases on the UK and in machine readable form, are on this disc and a sample of the list includes the 1981 Census, the General Household and the Family Expenditure Surveys, the BBC's survey of Daily Life in the 1980s, species maps, SSSI conservation sites, The Mortality Atlas, and industry location data. In addition, hundreds of tabular data sets in non-machine readable form (NMR) were converted for inclusion on this disc. Reformatting data sets from various host mainframe computers and converting NMR data sets required considerable computer and people power. The end result is a statistical database with a wide range of applications for the educational, institutional and commercial sectors.

Depending on the means of data collection, the data can be displayed either as bar/pie charts or as mappable data sets. The facility for mapping statistical data held in digital form over a map base has up

to now only been available to geographers using a million pound mainframe computer and specialist software packages. In effect, with the Domesday system the user can select an area in the UK from over 500 specially redrawn OS maps and have statistical data overlaid in colour coded kilometre or raster grids. It is then possible to analyse the data in greater detail, to reclassify the information using statistical routines built into the software and to retrieve names of area types and values for particular grid squares.

Anyone wishing to undertake research on any one of a wide variety of topics can ascertain whether the statistical research has already been done by checking Domesday first. It is also possible to determine where further information can be obtained by researching the Domesday sources. Each data set has related text pages which provides information about the source, the means of data collection and the copyright holder.

Finally, in order to include information on life in the 1980s which could not be represented statistically or pictorially, a special editorial team researched and selected over 2,000 previously published articles from newspapers, magazines, journals, *Hansard* and books. In addition, 45 national subject experts wrote specially commissioned articles to provide an overview on topics such as communications, defence, health, the social services, sport and education.

Access to the half a million screens of information on the National Disc is by three means: 'Gallery', a walk through the pictorial elements of the disc; 'Contents', a hierarchical information structure; and 'Find', a keyworded index system with over 6,000 stem keywords. As on the Community Disc access to information is by selection with the tracker ball or keyboard from choices offered by the menu bars or options.

BBC ADVANCED INTERACTIVE VIDEO TECHNOLOGY

In order to store and retrieve the vast amounts of video and digital information on the two Domesday discs, the BBC, Philips and Acorn have developed the BBC Advanced Interactive Video system. The AIV system includes a BBC Master Series microcomputer from Acorn, a BBC AIV LaserVision ROM player from Philips, and an RGB analogue monitor and a tracker ball control. The whole system will play other LaserVision discs and other interactive video materials with floppy disc software for BBC computers. However, the Domesday Discs can only be played on this AIV system as no other interactive video system player has the capability of reading analogue and digital



Phyllis S. Gove with one of the Domesday discs.

information from an optical LV ROM disc. The Domesday retrieval and indexing software, developed by Logica Software for the BBC, is stored on the optical disc. Each disc side is capable of storing 54,000 still frame photographs and 324 megabytes of digital information.

Applications and Resources

As the first system to accommodate and allow easy access to a wide range of video and digital information, Domesday is proving of value to users in schools, colleges and universities, libraries, commercial organisations, courier and distribution services, local and national government offices, land and estate agencies, journalism, film and television companies, regional development agencies, tourism and travel organisations, and of course to the general public.

Groups of teachers throughout the UK working in collaboration with BBC Enterprises have written booklets on how to use the Domesday discs as a resource in the following curricular areas: mathematics; geography; art, craft design and technology; social sciences; primary projects and topics; and for in-service training.

The first AIV disc to follow Domesday, the 'Ecodisc', was launched in May 1987 and allows easy access to a database in moving and still pictures, text and statistical information based on a nature reserve in the south west of England. Users are able to see the effect of their decisions taken in managing the reserve, study the environment in different seasons, take a surrogate walk through the reserve, and conduct a range of field studies. Other AIV discs are being planned, and BBC secondary computer software programs for use with The Domesday Discs were launched in September 1987 along with the seven resource booklets mentioned above.

Differing substantially from the two Domesday Discs, The Ecodisc simulates changes in the environment over time and as a result of actions taken by the user. Although developed by BBC schools TV Producer, Peter Bratt, for secondary level students aged 15 to 17, initial trials have already indicated a much wider range of users from primary level pupils to management trainees in industry.

Conclusion

The BBC has undertaken the major task of co-ordinating data collection on an enormous scale and working in co-operation with other partners to develop the AIV hardware. Other countries are presently

investigating the feasibility of conducting Domesday equivalent projects, and other groups working with the BBC are developing database video-discs using the AIV format. A new era for interactive video has begun with Domesday as an important advance in electronic storage, publishing, and information retrieval.

Phyllis S. Gove is currently responsible for co-ordinating the production and the editing of a series of resource booklets on how to use the Domesday Discs in the curriculum, and providing consultancy services to fledgling Domesday equivalent projects. She was previously Education Officer on the Domesday Project and responsible for co-ordinating the Domesday data collection process by the 15,500 schools and community groups in 1985-86.

THE NORTH WEST EDUCATIONAL COMPUTING PROJECT

A CASE STUDY

Ian Robertson and Mike Picciotto

1. BACKGROUND

The Interactive Learning Unit at Lancashire Polytechnic has been set up to exploit and develop the research and development work done by the North West Educational Computing Project. This was a project supported by an academic/industrial partnership to investigate the potential of interactive video for the teaching of foreign languages. The project ran from January 1985 to July 31st 1987.

The three main collaborating partners in the North West Project were:

- Lancashire Polytechnic
- IBM United Kingdom Trust
- the University of Manchester

The project also received support from four Local Education Authorities in the North West, each of which nominated a school to participate in the project, so providing four testbeds for the piloting, trialling and evaluation of systems and materials.

2. OBJECTIVES AND TARGET GROUPS

The objective of the project was to develop a suite of software capable of exploiting a configuration comprising an IBM Personal Computer and a laser disc player together with associated systems and applications software. In practical terms, the project team's task was to research interactive video in language learning through the design, implementation, trialling and evaluation of working educational programmes.

The target groups were young language learners aged 11-16 in four trial schools in the North West of England. As part of their participation in the project, they received donated IBM computer equipment

and software, training in the use of generic software, authored language learning programmes produced by the project and finally, interactive video workstations and courseware.

3. THE DELIVERY SYSTEM

The delivery system comprises one IBM PC, one Sony KX 14CPI monitor and one Philips VP835/831/410 laser disc player.

Installed in the IBM PC is a MIC 2000 controller board which serves as an interface between the computer and video player. The board synchronises computer and video signals into a single picture, allows for independent software control of the separate video, audio and RGB signals, and makes the system PAL/NTSC transparent.

4. NWECP INTERACTIVE VIDEO COURSEWARE

In the past, educational computing in the area of modern languages has been limited to drill and practice of language which had already been taught. In the UK there has been considerable resistance to computer use on the part of foreign language teachers, which may at least partly be because conventional computer-assisted techniques are seen as inappropriate for communicative language learning.

The aim of the team has from the start been to design materials which would be unique in their capability of teaching foreign languages at the workstation.

All the work provided for learners is oral and in the foreign language. Interaction with the videodisc is controlled by single key presses, and no text input is ever required.

It is estimated that the completed research and development has resulted in at least 100 hours of work for an individual learner or small group.

4.1 TEMPLATE 1

The programs of the first Template are structured language acquisition programs based on the bilingual method originated by Professor *Carl Dodson* of the University College of Wales and using available off-the-shelf videodiscs.

The teaching programs which make up Template 1 are based on the 24 main dialogues of the BBC's Ensemble series which is available on two double-sided active play videodiscs.

Each dialogue is presented in three teaching stages, or 'steps'.

Step 1 Imitation

The whole dialogue is first presented without support.

Each dialogue is then divided into utterances of manageable length. The learner imitates the foreign language utterances by speaking them aloud. There is support from foreign language and native language subtitles.

Learners may repeat each utterance at will, continue to the next utterance or return to the previous utterance allowing complete freedom of movement through the dialogues.

Learners may decide for themselves how often to repeat a particular utterance and when to move on to the next.

Native language support is displayed every fourth repetition.

During step 1 users learn to say and understand the foreign language utterances.

Users may exit from the programme at any time by pressing <Escape>.

Step 2 Interpretation

The learner sees the same segments, but this time silenced and with the native language subtitle. The learner must speak the French equivalent.

Audio feedback is available by pressing the F9 key. If French text is also required it is available by pressing the + key.

Both these keys are repeat keys, allowing learners the opportunity for more practice if required. Just as in step 1, learners may repeat, continue and return at will.

In step 2 the links between the two language systems are forged and consolidated.

Once again, the escape key allows users to exit from the programme at any time.

Step 3 Substitution and Extension

The elements of language acquired in Steps 1 and 2 are permuted to give many variations, and are extended to produce longer utterances. This is done in the form of substitution tables.

A single key press begins a randomising procedure, a press of the



Mike Picciotto and Ian Robertson.

space bar selects a random element from each column and displays the utterance in a box in the lower half of the screen. The learner speaks the foreign language equivalent and a second press of the space bar confirms the answer. Another press of F9 restarts the process.

When sufficient work has been done on one table, the next can be selected by pressing the + key.

In this way, it is possible to exploit the wider possibilities of the language contained in each dialogue.

It is estimated that the present Step 3 tables can display more than 100,000 separate cues and responses.

4.2 TEMPLATE 2

Template 2 programs, based on the Project's own videodiscs, are intended to provide a much more open-ended learning experience, allowing learners more freedom to shape their own learning strategies.

The language on the videodiscs is also natural speed, exploiting to a much greater extent the flexibility and responsiveness of the developed software. It is felt that Template 2 programs provide access

to learning experiences unavailable from either other technological media or human teachers.

The Dialogues

The dialogues were shot using amateur actors on location in schools in France and Germany. Brief scenarios were drawn up by members of the project team and sent to contact schools where dialogues were scripted and acted mostly by young people in the same age range as the target learners. In the course of this filming it was necessary to adopt some quite unconventional strategies to overcome the time constraints and the inexperience of the amateur actors. In all, twelve French and thirteen German dialogues were shot in eight days.

Viewing

In Template 2, learners are offered the facility to view the dialogues in a variety of ways — paused or unpaused, with or without sound, with foreign language subtitles, with English language subtitles, with both or with neither. In all, there are 16 possible ways to view each dialogue.

It is possible, therefore, to choose an unpaused viewing with sound but no subtitles in order to view the dialogue straight through. Unlike Template 1, however, this choice is not imposed on the learner.

To work intensively with the dialogue, learners would choose a paused viewing, perhaps with sound and with foreign language subtitles.

In a paused dialogue, the utterances are longer and faster than in Template 1, reflecting the more natural language employed in the dialogues. More highly developed software has however improved access times, so that utterances can be recalled instantaneously. Subtitling is erased before learners imitate the utterance, thereby preventing interference from the text. Available keys are: F9 to repeat, spacebar to continue to the next utterance, F1 to return to the previous one, + to repeat the current utterance with the opposite choice of subtitles and <Return> to toggle sound on or off.

For this option as for all others, the keys are chosen to fall easily under the hands by touch so that the learner's attention is not distracted from the screen. The use of the keyboard is considered more practical than a touch screen for this application in view of the intensive interaction involved. The remainder of the keyboard, i.e. all keys which are not 'live', is disabled, so that the only method of leaving

the programme is to press <Escape> which is available at all times. This robustness was considered essential for school use, but is also extremely comforting to adult users.

It has been found that, although the amateur actors are sometimes less convincing than professionals in unbroken sequences, they are often far more expressive in still frame. Template 2 dialogues are very powerful in exploiting the non-verbal and paralinguistic elements of communicative interactions.

Printing

A printing menu allows learners to print either the whole of a dialogue or each individual rôle separately.

Activities

Finally there is the activities menu which offers learners a range of activities with the dialogues. Learners or teachers can browse across the whole surface of the disc, employing most of the facilities of the videodisc player.

The most important learning option is the rôle-play. In this option, learners or groups of learners have the facility, having first had the opportunity to learn the language in the dialogue, to view each character in the chosen dialogue, to select one, and to engage in a rôle-play with characters on the videodisc.

Support available includes the ability to reveal the subtitle one word at a time, to view the silenced utterance or to repeat the utterance with or without sound an unlimited number of times.

There are also games designed to focus the attention of learners on non-verbal and paralinguistic clues in the communicative process, such as surroundings, body-language, facial expressions and so on. These games involve matching silent video sequences to one audio clip or audio clips to one silent video sequence.

The system can be adapted not only for the teaching of French and German but to any language which has an accompanying videodisc. The system could also be used to teach English from any other language!

4.3 THE STUDY OF LANGUAGE FUNCTIONS

Further experimental work has included a feature in which users can select from main and submenus of communicative functions and view

or work with all examples of a particular language use which are present on the videodisc. In this case the data retrieval facilities of the computer combined with the accuracy and speed of videodisc access can offer learning possibilities available neither from a human teacher nor any other form of educational technology.

5. EVALUATION

The first available evaluation of specifically designed educational IV programs in the UK has emerged from the four trial schools of the NWECP which began to receive workstations and software in November 1986. The evaluation centred on establishing the attitudes of users to the system by observation and interviews with pupils and teachers based on questionnaires.

5.1 PUPILS

Pupils have reacted very positively to being put in charge of their own learning and being allowed to develop their own strategies. They enjoy working in a more private situation, although the systems are used by small groups rather than by individuals as would be the case in industrial training.

5.2 TEACHERS

Teachers have reacted very enthusiastically to the fact that pupils can work at their own pace. They report a highly responsible attitude on the part of pupils, even when they use the equipment unsupervised, and even previously poorly motivated groups have been enthusiastic about using the systems. In one school teachers report pupils arriving well before 8.00 a.m. demanding entry to the language room to use the equipment, and in the same school the take-up rate for the second foreign language in a third year group which uses the IV system intensively has risen from the usual seven to twenty-one!

As a member of the North West Educational Computing Project team Ian Robertson was responsible for the design, development and implementation of system hardware and software.

Mike Picciotto, a language teacher for 19 years, has worked as project officer on the North West Educational Computing Project.

RECORDING THE ITALIAN CULTURAL HERITAGE

A CHALLENGING AREA FOR OPTICAL MEMORY SYSTEMS

Ernesto Bartolozzi

0. CONTENT

The paper gives a brief description of the products, based on Optical Memory Systems (OMS), that SIDAC will be implementing in the area of the Italian Cultural Heritage (ICH).

A list of the major questions to be answered is also given.

The paper is subdivided into six sections:

1. The area of interest
2. SIDAC OMS for the ICH
3. Two SIDAC OMS for the Italian National Library Holdings System (NLHS)
4. Two SIDAC OMS for the Italian National Photograph Archives (NPA)
5. Two major problems
6. A videomuseum

1. THE AREA OF INTEREST

In 1986 the Italian Ministries of Culture and Labour decided to dedicate significant financing to a number of projects related to:

- a) the management of the Italian Cultural Heritage (ICH)
- b) the improvement of its fruition both by scientists and by the (general) public

This was decided on the basis of two basic considerations:

- a) ICH is a real business for Italy (1.6% of GNP)
- b) the projects are to create important employment opportunities

Promoting the headline «Recovering the Italian Cultural Gold Mines» was a great success in a country, like Italy, which is virtually without raw materials.

As indicated by this fascinating slogan, the main objectives of the program were meant to deal with: discovery, description, storage, fruition on site (access, consultation), remote fruition (transmission, publishing) of all the historical, artistic, architectural and archeological treasures that survive today.

The most challenging aspect the project designers have to face is the size of ICH: nobody knows exactly how many items should be recorded, also because the boundaries of the area of interest are not generally defined (painting, sculpture, architecture, music, archeology, history, folklore, anthropology, ethnography, environment, territory ...); the most credited estimates span from some 50 to 100 million items! And, for the moment, only about 20 million items are supposed to be already catalogued!

2. SIDAC OMS FOR ICH

Recording ICH appears to be a very promising area for OMS (Videodisc, CD-ROM, WORM ...): in fact ICH concerns a very large amount of pictures, texts and sounds, so that OMS can be used both for storage and fruition (both on site and remote).

As a matter of fact many of the projects approved by the Italian Ministries of Culture and Labour are based on OMS.

SIDAC, a System Integrator in the area of OMS, is a Leader Company in Italy, having already developed OMS programs in fields such as art («Leonardo da Vinci», «Galleria Spada»), tourism («Hotel Catalogue», «Gallipoli») medicine («Trachea», «Prostate Surgery», «Enteropathy»), training («The ABC of Informatics»). SIDAC is one of the partners in charge for two major projects already approved by the Ministries:

1. National Library Holdings System (NLHS)
2. National Photograph Archives (NPA)

The role of SIDAC in these projects is to implement OMS to support storage, on-site fruition and publishing of the project-related part of ICH. In fact, SIDAC will provide various kinds of OMS: Videodisc Systems driven by the Central Site, Videodisc plus CD-ROM Systems, purely textual CD-ROM systems, picture plus text CD-ROM systems ...

The two projects started on April 1st. They span over a period of three years. They are assumed to be the first step of a much longer activity.

Some of the products SIDAC is going to develop should give the opportunity to implement significant CD-I prototypes/applications.

3. TWO SIDAC OMS FOR THE ITALIAN NATIONAL LIBRARY HOLDINGS SYSTEM (NLHS)

The main objectives of NLHS are:

1. To record information about books and other written documentation of NLHS
2. To promote circulation of the information

The main functional requirements of the project for NLHS are:

1. To allow librarians to more easily manage the material of NLHS (National Librarian Index)
2. To improve public access to the information, in a safe way
3. To improve circulation and publishing of the information

Several million items (50 million?), very few of which already catalogued (50 thousand?), should be managed, in the long term, by NLHS.

SIDAC is going to implement two OMS for NLHS:

1. The «BNI/BOMS» Catalogue
2. The «Beneventane» Catalogue

«BNI/BOMS» is a bibliographic catalogue containing significant parts both of the Italian National Library (Biblioteca Nazionale Italiana — BNI) and the Bulletin of Modern Foreign Works (Bollettino delle Opere Moderne Straniere — BOMS). The entirely textual catalogue will be stored on CD-ROM; therefore the «stand-alone» workstation (see fig. 1) is configured as follows: a CD-ROM player, a personal computer, a laser printer.

Users will be able to consult the bibliography by a guided search system using single search words or combinations of words to specify the query.

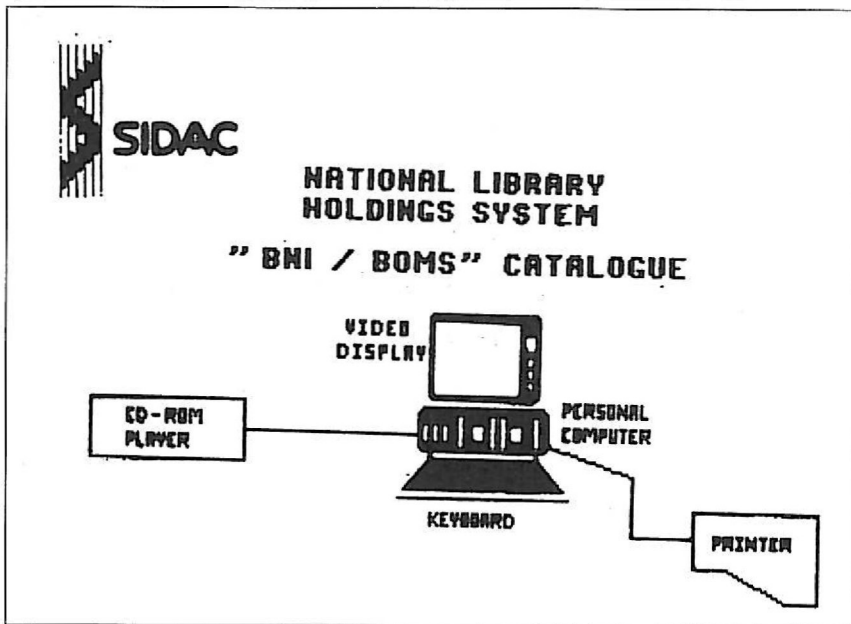


Fig. 1

4. TWO SIDAC OMS FOR THE ITALIAN NATIONAL PHOTOGRAPH ARCHIVES (NPA)

The main objectives of NPA are:

1. To record pictures and descriptions (about 2,000 characters per picture) of ICH
2. To promote circulation of information

The main functional requirements of the project for NPA are:

1. To allow librarians to more easily manage the material of NPA
2. To improve public access to the Photograph Archives through easy procedures
3. To improve circulation and publishing of the content of NPA
4. To build-up a «model» for the management and access to Photograph Archives to be used as a reference for other similar applications

About 3,5 million pictures, 500 thousand of which already catalogued, should be managed, in the long term, by NPA.

SIDAC is going to implement two OMS for NPA:

1. Main Site
2. Local Site and Specialized Catalogues

The configuration of «Main Site», to be installed at the Central Building of NPA, is (see fig. 2):

- a mainframe
- a data base containing the descriptions of the pictures
- the information retrieval software
- some videodiscs on which pictures are stored
- many terminal stations for data-entry (local stations)
- many terminal stations for enquiry (both local and remote stations)

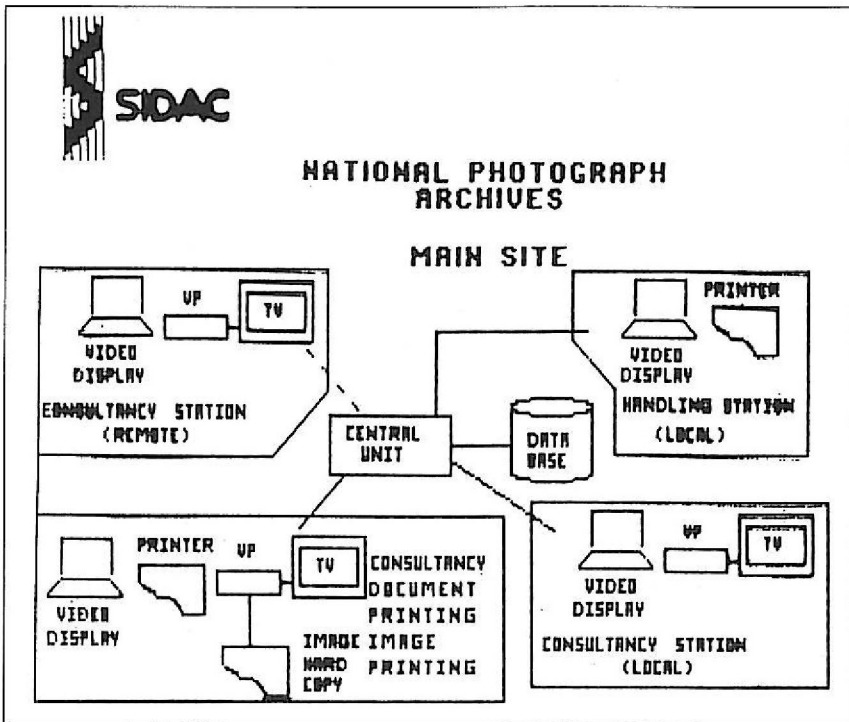


Fig. 2

- many terminal stations for photos, displays and copy (both local and remote stations)

«Beneventane» is a catalogue of handwritten medieval music manuscripts. Each page is described and identified with textual data of about 500 characters. Pictures and texts will be stored on CD-ROM: the pictures will be displayed in black and white.

The implementation of the «Beneventane» catalogue will be a very delicate matter, because:

- documents cannot be taken away from the museum
- they are very delicate
- they are of variable size (e.g. 155x95 mm, 300x220 mm, 420x300 mm)
- they are not always easily readable
- text and notations are often brilliantly coloured while the background is almost always pale yellow or pale brown
- notations are often very small, compared with the text, and placed at the bottom or aside of the page

The «stand-alone» station includes (see fig. 3): a personal computer

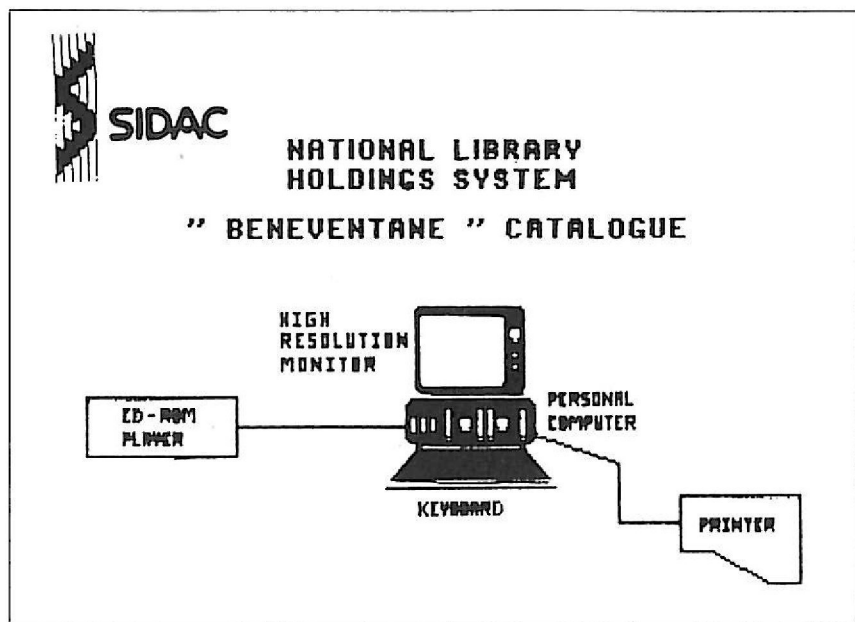


Fig. 3

with high resolution monitor and graphic capability, a CD-ROM player, a laser printer.

The software will allow information retrieval, «zooming» and/or side-way scrolling, hard copy on a laser printer.

There are some open questions to be answered in the very near future:

- Which level of resolution could be economically acceptable both for scanner and display?
- What about the use of CD-I?
- And what about the use of Laserdisc, with «false» zooming?
- Could LV-ROM be used?

The configuration of «Local Site and Specialised Catalogues», to be installed at some peripheral locations is (see fig. 4):

- Personal Computer
- I.R. software package to handle simultaneously CD-ROM and videodisc
- CD-ROM player
- Videodisc player

The «Local Site» contains the same information (pictures and texts)

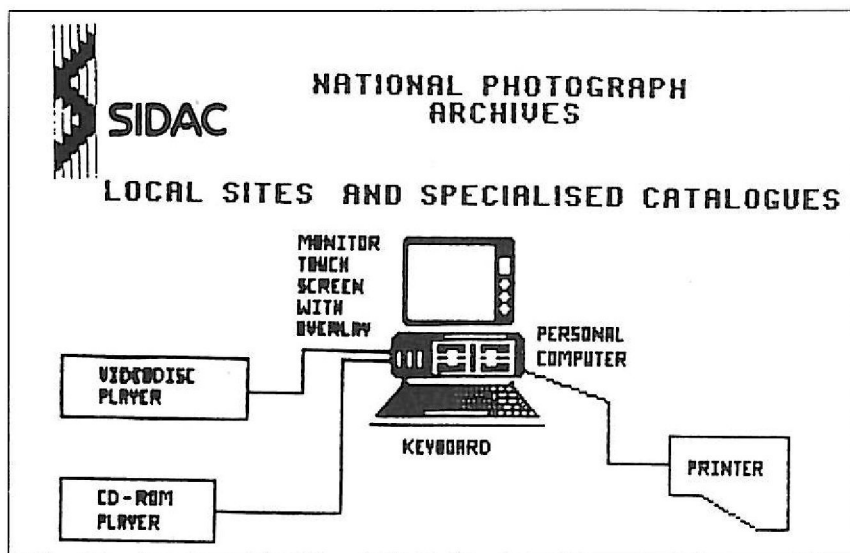


Fig. 4

as the «Main Site», the «specialised catalogues» contain a sub-set of the main catalogue.

There are some open questions to be answered in the very near future:

- How can the system be designed so as to be capable of handling archives expansions? (a juke-box could be used, both for main and local site)
- How Laserdisc and CD-ROM technologies are to converge towards CD-I?
- How economically feasible is the use of a LV-ROM instead of Laserdisc plus CD-ROM for the «Local Site»?
- Once the system has been built up, would it be possible to connect other systems configured in the same way?
- How can a user access various image and data bases without buying them all?
- How economically feasible is image digitalization and transmission for black and white photos?
- Is it possible to have the same image bank accessed simultaneously by many users? (i.e. video display with screen memory)

5. TWO MAJOR PROBLEMS

The two major problems that should be always born in mind all along implementation of OMS for ICH are:

1. The size of the problem
2. The definition and evolution of the «standards» of OMS

6. A CD-I PROTOTYPE

As outlined in par. 2, SIDAC is going to implement some CD-I prototype applications in the area of ICH.

The first prototype which will be developed in cooperation with Philips, will be related to the «Beneventane» Catalogue: this application seems to open a very interesting opportunity to implement a real multimedia information product, containing images (the musical manuscripts), text (the description of the manuscripts) and sound (the oral interpretation of music).

7. A VIDEOMUSEUM

In collaboration with the Italian Ministry of Culture SIDAC has produced one of the most fascinating applications for interactive videodisc systems introducing art and culture to the public by transferring museum and art collections onto optical memories. We call it a videomuseum.

A videomuseum is far more than an audio-visual. It simulates an actual museum «visit» and it is always available even during closing hours!

The visitor can choose to take a quick look at the grounds, or browse back and forth at will, or stop to contemplate some exquisite detail of his favorite masterpiece, or he can join a guided tour (in a choice of languages) through the gallery halls.

If he's a student or a scholar, or simply interested in further knowledge, he can refer to the Videomuseum Data Bank for all the information he needs on the museum, such as its history and architecture, technical and historical notes on the masterpieces, or the artists' biographies.

SIDAC is proud to present «Galleria Spada», the first videomuseum on optical memories in Italy, on which the magnificent Spada Palace in Rome, and its treasures, are recorded.

The Spada Gallery is one of the few and most important examples of a museum originated by an ancient private collection, still intact today.

Palazzo Spada is the historic residence of the Spada family, of whose illustrious members Cardinal Bernardino Spada (1594-1661) was the greatest art patron.

The many paintings collected by the Cardinal during his appointment as Apostolic Ambassador to Bologna (1627-1631), belonging mainly to the local Emilia and Romagna schools, brought new grace to the grand 12th century paintings that had dominated the family collection until that time.

Bernardino's nephew, Cardinal Fabrizio Spada Veralli, and other heirs continued to enrich the collection throughout the 17th and 18th centuries, with paintings and sculptures acquired as inheritance, dowries, or purchases.

The advent of the French Revolution brought a halt to the collection, which lasted throughout the 19th century.

The Italian Government acquired the Spada Palace with all its belongings in 1927, and the collection was opened to the public. After the Second World War, the Spada Palace was restored. New paintings



From the «Galleria Spada» videodisc. The painting shown in the bottom left-hand corner of the top frame is shown as a close-up in the bottom frame.

were bought to replace some of those lost during the war, and finally the museum was once again opened to the public.

The Galleria Spada Videomuseum, produced by SIDAC in cooperation with the Italian Ministry of Culture, contains over 5,000 still pictures plus film sequences, and an image and data bank with catalogue information. The system is composed of a videodisc player, a personal computer, and a TV monitor, on which the visitor can select his choices on a touch screen.

Ernesto Bartolozzi is Director of SIDAC — Società Italiana per la Diffusione Automatizzata di Informazione Culturali e Tecniche p.A.

A CD-ROM-BASED GEOGRAPHIC INFORMATION SYSTEM

Erling Maartmann-Moe

1. INTRODUCTION

In April this year the first Norwegian CD-ROM was presented. It was the result of a cooperation project between four partners that started almost a year earlier. The participants are major information distributors in Norway, and the goal of the project was to gain experience with CD-ROM and CD-ROM production, and to evaluate the medium as a carrier for information, in particular geographic information. The Norwegian Computing Center, a research center for applied computer science, were project leaders, but also contributed to the contents of the disk and with software. The work was funded partially by a research grant aimed at investigating the suitability of optical media for storage and distribution of maps and geodata.

The participants in the project were:

- The Central Bureau of Statistics of Norway
- The State Mapping Agency of Norway
- Cappelen, a large publishing house
- the Norwegian Computing Center

Although the participants in principle were independent and free to put what information they wanted on the disk, it was evident that these partners possessed information that could be combined in a creative way to form a prototype for a CD-ROM-based geographic information system. But before I discuss the concept of Geographic Information Systems in more detail, I will describe the data from each participant.

2. POPULATION STATISTICS

The Central Bureau of Statistics is of course the main distributor of statistical data in Norway. In addition to a large number of statistical publications, they provide the state and local administrations with

statistical data in machine-readable form. Up to now these data have been distributed mainly on tapes, each tape made on specific order from the customer. Besides being a labour-intensive process, the tape format does not give the increasing number of PC users a practical way to process large volumes of statistical data. The Central Bureau of Statistics decided to prepare a total of 50 MB of population statistics for CD-ROM. This database contains detailed population statistics for all Norwegian municipalities for the years 1976 to 1986, and also population forecasts for the same municipalities for 35 years on until the year 2011. Printed in tabular form, this database would require approximately 8000 pages.

The user interface of the database is menu-oriented and gives the user a choice of seeing the data in 4 different tabular forms on the screen, or to use the files as input for other computer programs. It also contains a catalogue of municipality codes, and detailed file descriptions.

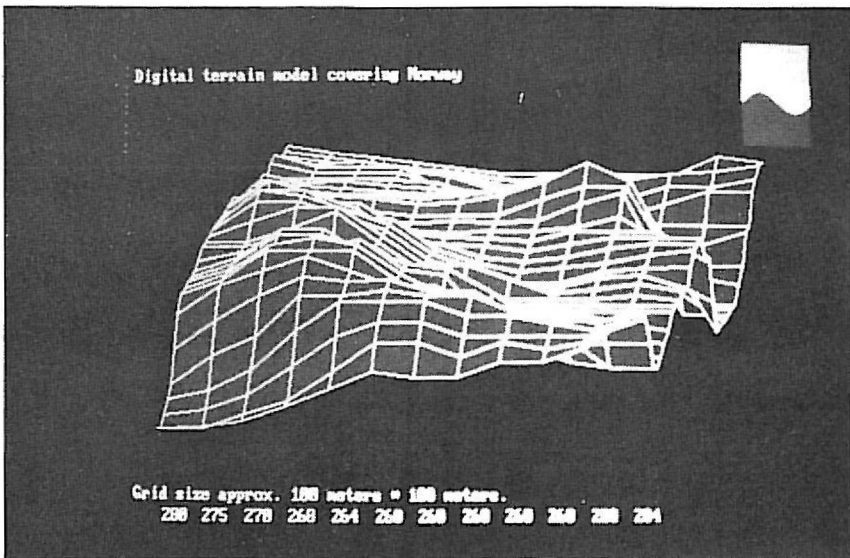
The data may be viewed either by year or by municipality. To compensate for the rather slow access on a CD-ROM, the tables were pre-processed and stored in both formats on the CD-ROM. This ensures quite good response times from the system even when data from several files are retrieved in one query.

3. MAPS AND SATELLITE IMAGES

The State Mapping Agency has for the last 10-15 years increasingly focused upon computer-based methods in mapping, and the distribution of maps and other geographic information in machine-readable form. CD-ROM is a possible carrier for relatively stable geographic information to several users. This new and compact format may be used in ships for navigation, in cars, by professional planners, by energy- and telecommunications administrations and other professional map users.

The contribution from the State Mapping Agency to the CD-ROM contains 24 files with map and image data from different areas in various scales. This totals a volume of about 26 MB. The data are of four types:

- Maps on vector format
- Maps in compressed raster format
- Satellite images in uncompressed raster format
- A database for a digital terrain model (DTM)



Excerpt from a Digital Terrain Model.

The whole of Norway in two different scales is represented. These are identical with the basic maps in the National Atlas of Norway.

From the principal map series M711 at a scale of 1:50,000 they have taken contour lines as well as drainage from selected areas. Two map sheets from the economic map series of Norway (1:5,000) are stored both in raster and vector format.

The vector data are stored in a format called SOSI, a symbolic format that is tree-structured. All data values are associated to names. The format allows the user to predefine names similar to the declaration of variables in programming.

The raster data are stored in a packed, run-length encoded format. The map is scanned with a resolution of 10 pixels/mm. The map sheet is subdivided in areas of 50x64 pixels, and this area is run-length encoded, giving a compression factor of about 8:1. An index table to these sub-areas of a map sheet is stored in the map file, and is kept in memory while the map is processed. The user will see only a part of the map sheet on the screen, the screen then composed of a number of sub-areas of the map. The screen acts as a window into the map. The reason for this sub-division is to facilitate updating of the screen when the user «navigates» in the map. This design makes updating

from CD-ROM as fast as using a hard disk on a PC.

A satellite image of the area where the State Mapping Agency is situated is also included. The image is taken from Landsat Thematic Mapper, and is stored as an uncompressed byte string corresponding to the screen it is displayed on.

The basis for a Digital Terrain Model (DTM) is a database of altitudes and point coordinates. From a 1:50,000 scale map the altitude in the points of a 2x2 mm grid is measured and stored together with the coordinates. This corresponds to one point for approximately every 90 meters in the terrain. The whole of Norway is digitized in this way, and this database would almost fill up a CD-ROM. On this disk coordinates for one map sheet are stored. DTMs are useful for the planning of roads and buildings, for analyzing aerial photographs, for navigation, and for military purposes. Today this database is stored on tapes, and is subdivided according to map sheets. This causes practical problems for the State Mapping Agency — the users often require parts of the database crossing map sheets, and distributing this DTM on CD-ROM would be very practical.

The user manual for the Norwegian register of real estate, addresses and buildings (GAB) is also included on the disk.

4. ENCYCLOPEDIA

The **Cappelen** publishing house has contributed with the 1987 supplementary volume of their encyclopedia, the book version released only a month before the CD-ROM. The encyclopedia is used through a powerful text retrieval system called TMS-Research. Since such a system through its word search facility relieves one from the alphabetical ordering necessary in a book, the encyclopedia is completely reorganized in thematic chapters. These chapters are furthermore hierarchically organized. Within each chapter the articles are in alphabetical order. This means that browsing in the encyclopedia becomes more meaningful. For example, after reading an article about the African country *Angola*, you may browse on and see articles about other African countries instead of *angora*, *angostura* and other irrelevant items.

Since it is a supplementary volume it mainly contains update articles relative to the main encyclopedia released 7-8 years ago. This means that the emphasis is quite heavily on geographic information in a wide sense — including articles about countries of the world and one article about each Norwegian municipality. So the encyclopedia articles add to the geographic perspective of this CD-ROM.



The 1987 supplementary volume of Cappelen's encyclopedia on CD-ROM.

5. A GEOGRAPHIC INFORMATION SYSTEM

The Norwegian Computing Center has, in addition to being project leaders, provided the disk with maps, satellite images and software. Our institute has for 15 years been a leading milieu in Norway for automated mapping and remote sensing. One particular area of interest has been so-called Geographic Information Systems (GIS). Before I describe what we have done with this particular CD-ROM, let me try to present how we view the concept of GIS.

In our view a GIS is not simply a computer system presenting geographic information. A GIS should be more than a *lookup* system, where the system designer has prepared the answer to all possible questions that can be posed to the system. A GIS should be able to *process* and *analyze* geographic information. To do so it must contain more than mere pictures of maps. It must contain a topological structure, that is descriptions of map elements and their relations to each other. A GIS should integrate spatial data with tabular non-spatial data, such as for example administrative data. It should be able to produce cartographic and tabular reports of the results of retrievals

and manipulations. It should be able to identify different elements of a map, be able to answer questions about borders, distances and properties.

A system developed by the Norwegian Computing Center for the Soil Survey Unit in Zambia in 1986 illustrates some of these points. The system is dedicated to the problem of land evaluation. Land evaluation involves the integration of a number of physical characteristics (e.g. soil properties, climatic variables, topography, geology, geomorphology) in order to determine the suitability for growing certain crops in particular areas. The suitability is also dependent on the farming inputs used, such as fertilizer, drainage and irrigation. There may also be human factors that could influence the «suitability» in a broader sense, such as religious beliefs, eating habits, traditional farming practices and so on.

The system uses three types of information. Firstly a soil map that shows the distribution of different types of soils. Secondly, an agro-ecological map that partitions the land in zones having homogeneous climatic properties across its area. By overlaying these two maps, one obtains a set of maps, or geo-units, with distinct properties. Thirdly, the system contains a set of rules, a knowledge base, that defines the suitability for growing any crop on any soil. These rules are based on a FAO scheme, but are adapted for Zambian conditions by the Soil Survey Unit in Zambia. The system can give output either by tabulated results of the evaluation, or by map output on colour screen or plotter.

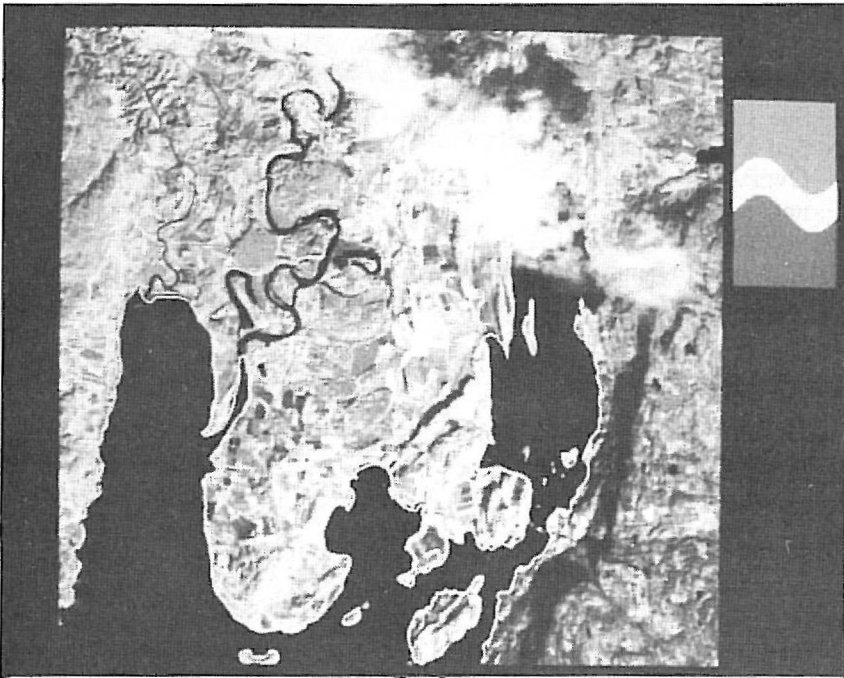
The system is designed as a tool for agricultural planning on the regional level. It is tailored to Zambian requirements, but is easily modifiable, and it is a good illustration of what a GIS is and can be used for. This Zambia system does not contain large amounts of data, and is therefore not included on the CD-ROM. A part of the software system that underlies the Zambia system is used, however, with a sample of Norwegian maps and population statistics.

On the disk is, in addition to the software itself, a detailed map of one county of Norway, and statistics for this area. The statistics are based on the 1980 census, and include variables like population, income distribution, housing, the number of children and elderly and so on. The geographical data are organized in a hierarchical topological structure, and can be shown as 4 different levels: census tract, enumeration district, municipality and county. Maps on any level are based on the most detailed level: The census tract. All statistical data are associated to this most detailed level, but they may also be aggre-

gated and shown on higher levels. The system gives results as a map shown on the screen with the result of the analysis shown as numbers, boxes, or shading.

On the disk some data redundancy is introduced to compensate for the slow access of the CD-ROM: If the user wants to see a map on the municipality level, a file that contains all the border elements of this municipality is accessed, instead of scanning one or several files which contain all the data down to the census tract level, and picking the border elements for the municipality. Each border element contains information about the area on its right and left side in addition to the (x,y)-coordinates constituting the border element. This is a flexible design that lets the user independently choose the scale of map and data, and also gives flexibility in presentation.

There is also a satellite image in two resolutions of the Oslo area on the disk, with masks for water, populated areas and so on.



A satellite image from Landsat Thematic-Mapper, overlaid with roads, rivers and outlines of densely populated areas.

6. CONCLUSION

I have now described the contents of the disk in some detail, and tried to give an idea of our conception of geographic information systems. These 4 sets of data, seen together and integrated in a common interface, constitutes a GIS over Norwegian municipalities. We think this would be a powerful tool for state and municipal planning, for education and for marketing people.

The 50 MB of statistical data on this disk does not cover all the parameters necessary in such a system. The CD-ROM with its large storage capacity and relatively low price both in hardware and software, is an ideal distribution medium for the data volumes that are needed for a system covering the whole of Norway in detail.

Since our disk does not cover the whole country in detail, it must be considered as a prototype to such a geographic information system. The basic GIS software is developed, but the different data sets on the disk are not fully integrated. In a fully developed system, the user could, for a specific area, e.g. a municipality, perform the following functions:

- See a picture of the town
- See the town's coat-of-arms
- See a satellite image of the area
- Overlay the satellite image with maps of roads, housing, borders and so on.
- See a map of the area at a specified scale
- Switch to the encyclopedia and get an article describing the place
- Get a list of all persons in the encyclopedia associated in some way to the place
- See statistics in the form of tables over population, income, elections, land use and so on
- Get the same statistics displayed as thematic maps, with a choice of resolution (municipality, enumeration district, census tract) on both map and data, and a choice of presentation of data (numbers, shading, colours, boxes).
- Perform statistical analysis, and also integrate data supplied on other media than CD-ROM

Such a system would obviously be of great interest in schools, given the possibility to display and analyze information about all parts of the country. But it would also be a valuable tool for planners in deciding where and when to build schools, homes for the elderly and

kindergartens. It may be used as a tool in planning public transportation. To marketing and information people it provides a tool to find where the potential market for specific goods and services is located, and how to distribute information to that area. This list of applications is easily expandable.

We are currently having discussions with the school and the mapping authorities in Norway about making such a system.

BIBLIOGRAPHY

- Maartmann-Moe, Erling: Optical Storage in Geographic Applications. TOC Conference on Optical Data Storage. London February 1987.
- Robb, Margaret: Raster Geographic Information Systems. State of the Art. Norwegian Computing Center 1986.
- Robb, Margaret, Strand, Geir Harald and Wang, Bo: Land Evaluation System for Zambia: System Documentation and user's guide. Norwegian Computing Center 1986.
- Strand, Geir Harald: Efficient distribution of data to local planners. 12th Urban Data Management Systems. Blois, France May 1987.

Reprinted with kind permission from Eikon SpA.

Erling Maartmann-Moe is a researcher at the Norwegian Computing Center, Oslo.

FRAMANDSPRÅKUNDERVISNING OG DATA

OM INNFORING AV NYTT VERKTØY I FRAMANDSPRÅKUNDERVISNINGA

Bjørn Dommarsnes

FØRESETNADER

Spesielt

Maskinvare: Ved sentral styring gjennom Datasekretariatet unngjekk Noreg kaos i maskinparken då data skulle prøvast ut i skuleverket, men DS styrte diverre grunnskulen mot den, allereie då gamaldagse og svake 8 bits Tikimaskina. Ved Gol ungdomsskule las me maskin-spesifikasjonar og satsa rett: Scandis med fargeskjerm over heile lina. Det viser seg at fleire og fleire kjem til same konklusjon som Gol: Dess lenger du kjem ned i skulen, di sterkare maskin treng du, og desto viktigare vert fargar. Diverre har svært få grunnskular utstyr som fangar opp desse krava.

Programvare: Dei fleste skular har vel den erfaringa at det er satsa lite på program til språkundervisning. Både leverandørar og forlag verkar uinteresserte og fantasilause. Her har Gol ungdomsskule vore svært heldige i det at me har hatt ein rettleiar og programmerar med interesse for språk og data. Program utvikla ved Gol ungdomsskule av *Halvor Graver* har vore, og er, grunnstamma i den datastøtta framandspråkundervisninga vår.

Økonomi: Gol ungdomsskule vart i 1985 forsøkskule i data med ein barneskule, vidaregåande skular og ein spesialskule knytta til seg i Golprosjektet. Økonomien har difor tillete brukbart maskin- og programvareinnkjøp, betaling for ekstra arbeid i samband med data, og bra kursing av lærarar.

Organisering: Me hadde 12 Scandis på *eitt* datarom, og 2 lærarmaskiner (den eine også brukt i fysikk/kjemi/teknologi) på eit arbeidsrom.

No har me 9 Scandis på eit datarom, 3 Scandis på tilstøytande grupperom saman med 1 Tiki (16 bits) og 1 Spectravideo. Dette i kombinasjon med valfag maskinskriving. På arbeidsrommet har me 2 Scandis på trillebord (populærvogn frå Norlag a/s).

Felles

Kollegium: Me er vanlege moderat entusiastiske, *brente* lærarar, opptekne med problem og oppgåver inna heilt andre områder i skulen. Me er i tillegg bundne i frivillig lagsarbeid, politisk arbeid, usynleg hus- og skulearbeid, husbygging, private problem osb. Kort sagt: Me er vanlege menneske med ulike grader av tastatur- og teknikkskrekk.

Rom/skule: Gol ungdomsskule er ein tradisjonell bussmodellskule med romproblem.

INNFØRING

Oppstartinga vart mykje forseinka av leveringsproblem på Scandis pluss ein del maskineil saman med dårleg reparasjonskapasitet hjå leverandør. Me var òg innom ein dårleg harddisk eit halvår, noko som forseinka oss mykje og tok ein del pepp ut av forsøket. Harddisken fekk me bytt inn i diskettstasjonar, og fekk slik betre stabilitet. Dette skapte i neste omgang meirarbeid for oss med kopiering og oppdatering. Dessutan vert det på sikt dyrare, pluss at alle diskettane lagar arbeid med lagring og systematisering.

I forkant av, og under denne perioden, starta me utviklinga av dei programma for språkundervisning me har i dag. Grunnen var at me fann Profitekst og Wordstar lite eigna til det *me* hadde førestilt var datastøtta framandspråkundervisning. Me syntest òg dei var lite eigna som *introduksjonsprogram i data*, og altfor lite tilgjengelege til å vera *alle* elevane sitt første møte med ein *tekstbehandlar*. Me trudde og håpa utifrå dei signala me fekk at det ville koma ein menystyrt og lett tekstbehandlar, noko me fekk med Harmoni 1 og 2. Dette viste oss at det også kan vera nyttig å følgja nøye med og venta på det rette.

Då me så hadde våre eigne verktøyprogram, kursa og oppmuntra me språklærarane til å bruka og eksperimentera med programma. I byrjinga vart det ikkje laga bestemte mål for bruk av data i språkundervisninga med omsyn til bruksfrekvens og bruksmåte. Kvar lærar måtte prøva å innpassa data i si eiga undervisning. Seinare har me sett at minimumsmål må vera knytta til ein lokal læreplan, og det er dette me jobbar med no.

Programvare

Programvara vår er utvikla etter behov hjå *både* lærar og elev. Lærarane hadde behov for å kunna laga prøvar frå stoffet elevane jobba med, medan elevane hadde behov for å kunna øva inn ord og tekst på ein rask og effektiv måte. Slik vart Cloze-test, Gloze-test, Verbttest og Lesetrening utvikla. Krumtappen vår, Halvor Graver, vurderte våre idear utifrå kva ei datamaskin kan gjera, og kom sjølv med framlegg om kva me trengde. Krava til *tilgjengelegheit* var store, så me sit no med ei programpakke i framandspråk som har autostart, meny og ryddige skjermbilete der du *alltid* veit kvar du er i programmet, og du *alltid* kjem attende til menyen med ESC. Skjermbileta inneheld i tillegg skjulte hjelpetekstar. Alle programma er verktøyprogram, dvs. at læraren/kontorpersonalet/elevar skriv inn det stoffet/dei tekstane som skal brukast. Ein del av dataarbeidet vårt i språk har vore å lata elevane skriva inn tekstane. Til slutt har læraren samla det heile på *ein* diskett og laga classesett.

Programpakka inviterar til fornuftig bruk av data i undervisninga, men sjølv om den er god, manglar me ein komponent. Det er eit omsetjing/oppgåveprogram som ligg på same nivå. Det finst i Danmark, men berre for fransk, på Piccoline-maskina. Dette programmet er utvikla av *Lis Kornum* og heiter «Un Menu Française». Det bør snart koma på norske forlag for engelsk og tysk også.

Cloze-test

Dette programmet øver opp/testar generell språkdugleik. Læraren kan skriva inn deler av, eller heile læreboka på eitt eller fleire trinn. Tekstane kan så behandlast på to måtar. Felles for dei er at ord vert borte frå teksten, og oppgåva vert då å finna ut kva ord som manglar. Orda som vert borte, vert sletta vilkårlig etter innstillbar frekvens (t.d. kvart 3. eller kvart 5. ...). Slettinga kan òg gå etter ei ordliste på 40 ord, og i denne ordlista kan du sjølv skriva inn og endra som du vil. T.d. alle formene av *to be*, alle personlege pronomen osv., slik at oppgåva går på meir spesifikke grammatiske områder.

Ulempen med programmet er at det sjølvsagt ikkje går an å leggja inn alternativ/synonym til orda som vert sletta. Eleven må ved ei skjermoppgåve finna det eine sletta ordet. Difor er det lagt inn ei hjelp: eleven kan få første bokstaven i ordet dersom læraren stiller programmet slik.

MENY.

- A = Bruksanvisning.
- B = Oppgaver på skjerm.
- C = Oppgaver på skriver.
- D = Innskriving/retting/sletting av tekst.
- E = Div. større slettinger i kartoteket.
- F = Kopiering til annen diskett. (2 diskettstasjoner)
- G = Oppgavetype og bokstaver

VELG BOKSTAV:

Systemet kan avsluttes her med et trykk på ESC-tasten.

ESC-tasten vil i ALLE andre program gi denne siden.

Hovedmeny på Cloze-test.

Læreren kan ta ut cloze-testar eller meir spesifikke grammatikkoppgåver på papir. Her kan ein sjølvstgt ta meir omsyn til synonym ved å laga ein fyldig papirfasit med mange synonym. Det viktige er at læreren raskt kan laga mange varierte oppgåvar frå kjent eller ukjent tekst.

Cloze-test/luketest kan sjølvstgt brukast i morskåopplæringa også.

Glose-test

Dette programmet kan brukast til å øva ord, uttrykk og setningar. Programmet kan forholda seg til både bokmål og nynorsk samstundes, begge målformene med mange synonym. Ein kan gå frå framandspråket til bokmål/nynorsk og omvendt.

Ein kan laga papirprøvar her også; frå alle glosene, eit avgrensa antal stykker eller berre frå *eitt* stykke.

Verb-test

Dette programmet er laga spesielt for engelske verb (tre former), men kan òg brukast for tyske verb. Papirprøvar her og, med same valmogleheitene som på Glose-test.

1. MEET JOHN AND BILL

It was half past eight in the morning. John was walking to school for the last time before the summer holidays. It was quite a long way from his home to the school, but John took his time. He didn't hurry. He <1> that this was the last <2> of the school year.

When he <3> to school, the bell was already ringing. <4> ran up to the school entrance. <5> the door was closed. The other pupils <6> already singing in the hall. He was <7> again. After a few minutes a prefect <8> out of the school and asked him what <9> was doing and what his <10> was.

<Klasse : 1 Stykke : 1 Linje : 1- 13 av 13 linjer.

Skriv nr, svar
og trykk ↵
(Eks : 12, pupil)
2 forsøk: (2-1 p).

Svar her :

10 spørsmål igjen av 10 CTRL

POENS: 0 CTRL

CTRL N = Nytt stykke CTRL G = Gi karakter

↑ = 1 linje opp
↓ = 1 linje ned
OO = forrige side
OOO = neste side
F = Første side
T = Til siste side
? = hjelp på av.

Valg **B** på Cloze-test. Oppgåva er ferdiglaga, og eleven skal svara på dei nummererte oppgåvene. Hjelpeteksten er synleg.

Lesetrening

Dette programmet gjer at eleven kan øva opp det visuelle minnet. Innlagde ord (opptil 2500 i ti hovudgrupper, kvar med fem undergrupper) vert plukka ut og viste fram. Eleven kan så sletta ordet for så å prøva og skriva det rett. Den andre varianten er at ordet forsvinn av seg sjølv etter ei på førehand innstilt tid (eitt sekund, to sekund osb.).

I dette programmet kan ein òg leggja inn setningar.

I tillegg til dette utviklar me no eit spesielt skrivesystem der eleven kjem rett inn i «skrivearket» når han startar programmet (som har autostart). Programmet har innstillingar for fargar på meny, bokstavar og bakgrunn på skrivearket og menyfelt, stor/lita skrift, zooming av tekst (forstørring 4 gonger), bruk av tegnsett (der du plukkar dei fleste tegna på Scandis inn på plass i teksten), bruk av ordliste med automatisk rettefunksjon og innleggjing direkte i ordlista av nye ord, bruk av ordliste der du kan plukka orda inn på «arket» og laga setningar, mogleg å laga oppgåver i matematikk der *alle* teikn kan brukast pluss at du kan kobla inn kalkulator. Dette programmet vil me no prøva ut i spesialundervisning.

NO OG FRAMOVER

Når me møter lærarar frå andre skular og bygder, føler me oss privilegerte med maskinvare og programvare (spesielt synest me synd på dei med 8-bits Tikimaskiner). Men sjølv om situasjonen kan synast dyster, så er det viktig å vera klar over at datastøtta framandspråkundervisning ikkje er nokon fast størrelse, korkje organisasjonsmessig eller når det gjeld innhald (her særleg dataprograminnhald). Maskina er, saman med programmet, ein reiskap på line med dei andre me nyttar dagleg. Ei dårleg maskin eller eit dårleg program kan utløysa mykje fin språkaktivitet. Kanskje vil nokre av dei mest effektive datatimane vera timar der elevane vert irriterte over manglande alternativ i Cloze-test eller dårleg synonymutval i Glose-test. Diskusjonane, på «målspråket», som spring ut av slike situasjonar, kan vera oppklarande og fine som utgangspunkt for vidare læring. Det gjev mykje fin sjølvtilitt å kunna «slå» maskina, eller forstå kva som kunne vore betre i eit dataprogram.

Dei som sit og ventar på MASKINA eller PROGRAMMET som skal løysa alle vanskar, bør nok stikka fingeren i jorda og byrja å jobba med det dei har, for å få igang språkhandling *framfor* skjermen. Mellom elevar, og mellom lærar og elev utifrå eit *behov* hjå eleven. Det «perfekte» program ville kanskje redusera språkaktiviteten til eit minimum (yes, no, ehm), det såkalla «grunt-level».

Når det gjeld organisering av datastøtta framandspråkundervisning, vil eg kommentera den slik:

- a) **Rom** — Etter dei erfaringane me har, er det mogleg at det vil vera lettare å ta data i bruk dersom maskinene er i kvart einskild klasserom, lette å ta til. Datarom er tungdrivne saker og truleg best egna til valfag. Ein modell kan vera å plassera billige PC'ar på datarommet til valfag, og reservera dei sterke og dyre maskinene til klasseromma.
- b) **Frekvens** — Bruksfrekvensen vil vera avhengig av lærar, årstrinn og klasse. Som med alle reiskap, må læraren vurdera effekten utifrå dei måla han/ho har i undervisninga. Når det gjeld dei meir drill- og hjelpeundervisningsprega programma, så ser det ut til at effekten av desse (og motivasjonen til å bruka dei) er størst i 7. og 8. klasse. På desse trinna er programma ganske «slitesterke». Cloze-test med ukjent tekst kjem best inn i 9. klasse. Det er mogleg at konsentrerte arbeidsperiodar (3 til 4 timar same veka) er betre enn at timane vert strødd utover ein lang tidsperiode med 1 time for veka.
- c) **Organisering av undervisning** — Med alle maskinene samla på *eitt* rom, vert organiseringa spesiell. Erfaringa vår viser at uthald ved maskina med *eitt* program varierer frå elev til elev. Som i andre læringsituasjonar kan nokre elevar sitja heile timen med *ein* lang, vanskeleg Cloze-test, medan andre har problem med forenkla tekstar (Basic-tekst). Her har læraren, som alltid, eit differensieringsproblem. Men med heile læreboka liggjande inne i programpakka vår, tekst og gloser, er det lett å skifta frå det lette til det vanskelege og omvendt. Intensiteten i arbeidet ved ei datamaskin tilseier likevel at arbeidsøkta bør vera noko mindre enn ein skuletime. Dette merkast betre når datamaskina har vorte ein alminneleg reiskap i språkundervisninga.

Med maskina i klasserommet kan dette ofte einsarta og langdryge opplegget lettare løysast opp med stasjonsundervisning. Sidan maskina er der i alle timar, kan enkeltelevar med spesielle behov lettare få meir maskintid.

Datamaskina har kome for å bli i norske klasserom. Om effekten kan vera både-og, og programvare og maskinvare så som så, ser ein lett alt no at det er ei elevgruppe som vil tena mykje på denne teknologien: elevar med ymse lærevanskar. På Gol jobbar med no med å utvikla ny programvare, og tilpassa gamal programvare til denne elevgruppa.

Cand.mag. Bjørn Dommarsnes arbeider i vikariat ved Glitrehaug spesialskule i Gol (permisjon frå Gol ungdomsskule).

DIALEKTEKSPERT

ET EKSPERTSYSTEM FOR DIALEKTGJENKJENNING

Torbjørn Nordgård

Sammendrag

I denne artikkelen presenteres en prototype til et ekspertsystem for dialektgjenkjenning. Det vil bli vist hvordan systemet *DIALEKTEKSPERT* resonnerer seg frem til hvilken dialekt en målprøve tilhører på basis av data som mates inn. Det blir også forklart hvordan systemet arbeider for å bevise hypoteser.

Første del tar for seg ekspertsystemer og ekspertsystemskall generelt. I del 2 beskrives *DIALEKTEKSPERT* på en uformell og ikke-teknisk måte. Del 3 går mer i detalj på hvordan systemet er implementert. Del 4 er en oppsummering og noen konkluderende betraktninger.

I. EKSPERTSYSTEMER OG EKSPERTSYSTEMSKALL

Ekspertsystemer er en forholdsvis ny subdisiplin innen informatikk. Den er nært knyttet til forskning omkring «Kunstig intelligens» (KI), og kan i mange tilfeller betraktes som «Anvendt KI». Et ekspertsystem har som mål å simulere ekspertkunnskaper. Det kommersielle potensiale for slike systemer er naturlig nok på områder der det er mangel på visse eksperter.

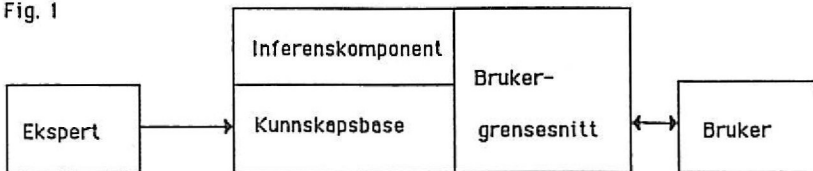
Det er ikke helt enkelt å trekke et skarpt skille mellom ekspertsystemer og mer «vanlige» dataprogrammer av en viss kompleksitet. En etter min mening god karakteristikkk av hva et ekspertsystem egentlig er, er gitt av British Computer Society's Specialist Group:⁽¹⁾

An expert system is regarded as the embodiment within a computer of a knowledgebased component, from an expert skill, in such a form that the system can offer intelligent advice or take an intelligent decision about a processing function. A desirable additional characteristic, which many would consider fundamental, is the capability of the system,

on demand, to justify its own line of reasoning in a manner directly intelligible to the enquirer. The style adopted to attain these characteristics is rulebased programming.

Et ekspertsystems organisering fremstilles ofte som

Fig. 1



En ekspert og en «kunnskapsingeniør» (dvs. en ekspertsystemkyndig person) bygger opp kunnskapsbasen som er et sett av regler som skal reflektere ekspertens innsikter. Inferenskomponenten (eller «slutningsmaskinen» om man vil) foretar logiske slutninger på grunnlag av kunnskapsbasen og de data som brukeren mater systemet med. Bruker-grensesnittet er brukerens kontakt/kommunikasjon med systemet. Det er mulig å kommunisere med ekspertsystemer på ulike måter, enten ved kommandoer som i «vanlige» operativsystemer eller ved (tilnærmet) naturlig språk. Dette varierer naturligvis fra system til system alt etter hvor høyt brukervennlighet er prioritert.

Et ekspertsystemskall er spesiallagde programmeringsomgivelser til å utvikle ekspertsystemer i, og hensikten er å gjøre det lettere å lage slike systemer. Dette er ofte ønskelig fordi ekspertsystemer er gjerne svært komplekse. Derfor er det viktig å ha tilgang til gode kontrollmekanismer slik at man lett kan få et inntrykk av hvordan systemet arbeider. Ekspertsystemskall inneholder predefinerte datatyper. Man er ikke nødt til å benytte disse i alle sammenhenger. Hvis det er behov for det kan man gjerne skrive egne småprogrammer og definere egne datatyper som ekspertsystemet kan bruke, samtidig som man holder seg innenfor rammen av et skall. Det er også mulig å ha ekspertsystemer innebygget i andre ekspertsystemer hvis det er ønskelig. Det viktigste ved et ekspertsystemskall er kanskje inferenskomponenten som systemutviklerne dermed slipper å skrive selv, men man må da innrette seg etter det regelformatet som forutsettes av slutningskomponenten og dermed også skallets begrensninger for hva som kan uttrykkes i reglene.

II. DIALEKTEKSPERT — EN UFORMELL BESKRIVELSE

Jeg har implementert en prototype til et ekspertsystem for dialektgjenkjenning — *DIALEKTEKSPERT*. Systemet benyttes på følgende måte: Brukeren gir det informasjon om målmerker,⁽²⁾ f.eks. om en dialektprøve har palatalisering. Etter hvert som målmerkene blir matet inn, vil systemet undersøke om målmerkene passer til ulike dialekter/dialektgrupper slik disse er beskrevet vha. reglene i kunnskapsbasen. Hvis systemet på bakgrunn av inndata finner ut at en målprøve kan sorteres under en gitt dialektgruppe, vil brukeren få beskjed om det (f.eks. om en målprøve tilhører gruppen av sørvestnorske dialekter). Denne prosessen vil foregå så lenge brukeren gidder å mate inn opplysninger om nye målmerker. Når systemet jobber på denne måten kalles resonneringsmetoden gjerne *forward chaining* eller *datadrevet resonnering*. Systemet trekker altså slutninger på bakgrunn av de opplysninger som tastes inn og gir beskjed om hva det finner ut. En annen resonneringsmetode er *backward chaining* eller *hypotesedrevet resonnering*. Da spør brukeren om en gitt målprøve tilhører en eller annen dialekt eller dialektgruppe. Systemets respons avhenger om det tillates å be om ekstra informasjon når det er nødvendig. La oss først anta at systemet ikke får lov til det. Brukeren vil da få enten benektende eller bekreftende svar avhengig av hvor mye informasjon som er matet inn om den aktuelle målprøven. Hvis det ikke er lagt inn noe målmerkeinformasjon vil svaret naturligvis være negativt. *DIALEKTEKSPERT* gir et bekreftende svar bare hvis målmerkeopplysningene for målprøven er konsistente med en eller flere regler i kunnskapsbasen for den aktuelle hypotesen.

Et eksempel: Systemet får beskjed om at i målprøve X er gammel-norske tostavingsord med kort rot redusert (altså «jamveksregelen»). Dessuten har historisk *a* utviklet seg til *e* i infinitiver og svake hunkjønnsord. Hvis brukeren nå spør om målprøve X tilhører dialektgruppen vestnorske *e*-mål vil systemet svare bekreftende. Hvis det derimot spørres om dialekten tilhører klassen av sørlandsdialekter, vil responsen være negativ.⁽³⁾

La oss nå anta at *DIALEKTEKSPERT* får lov til å be om tilleggsopplysninger. Hvis brukeren nå spør om den samme målprøve X tilhører klassen av sørlandsdialekter, vil systemet svare med et spørsmål, nemlig om svake hunkjønnsord i bestemt form entall ender på *a*. Hvis brukeren svarer benektende, vil hun få som svar at X ikke er noen sørlandsdialekt. Er svaret derimot bekreftende, må brukeren også svare på om målprøven har palatalisering. Hvis dette svaret er

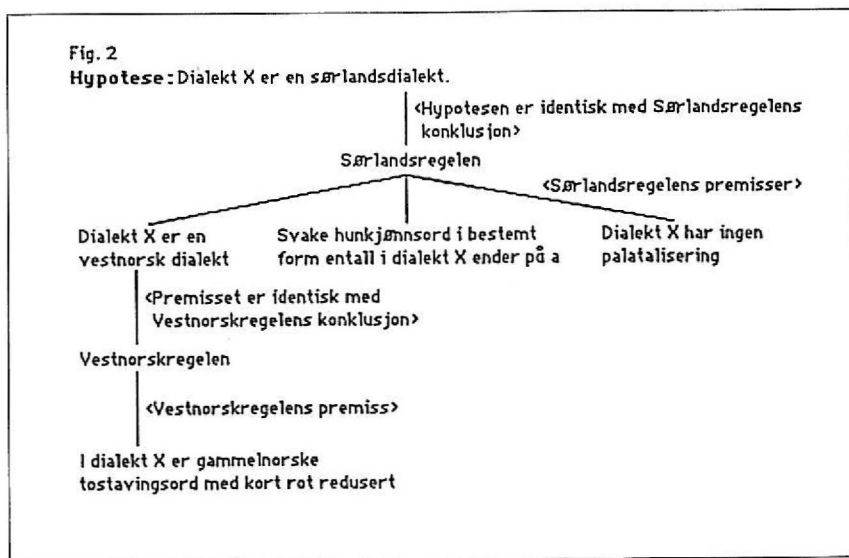
benektende, vil det endelige svaret være at X faktisk er en sørlandsdialekt. Merk spesielt at om man spør om en målprøve Y er en sørlandsdialekt uten at det er lagt inn noe informasjon om Y's målmerker og *DIALEKTEKSPERT* tillates å be om relevante opplysninger, vil brukeren bli spurt om målmerker helt til systemet kan avgjøre at Y ikke er en sørlandsdialekt eller at Y tilhører denne dialektgruppen.

La oss nå se litt nærmere på hvordan reglene ser ut og hvordan de blir benyttet i resonneringen. Alle reglene er på formen

HVIS premiss₁, premiss₂, ... premiss_n,

SÅ konklusjon

Et eksempel er *Vestnorskregelen* som sier at HVIS gammelnorske tostavingsord med kort rot er redusert i en gitt dialekt, SÅ er dialekten et medlem av klassen vestnorske dialekter. En regel kan naturligvis ha flere premisser, og et vesentlig poeng er at en regels konklusjon kan være et premiss i en annen regel. Det er bl.a. dette som gjør at systemet simulerer en intelligent adferd. *Sørlandsregelen* illustrerer dette på en instruktiv måte: HVIS X er en vestnorsk dialekt OG svake hunkjønnsord i bestemt form entall ender på a OG X ikke har palatalisering, SÅ er X medlem av klassen sørlandsdialekter. Dette kan visualiseres slik i en hypotesedrevet resonnering der hypotesen er at dialekt X er en sørlandsdialekt:



Systemet forsøker å finne en regel som har en konklusjon som er identisk med hypotesen. En passende regel blir funnet, nemlig *Sørlandsregelen*. Hvis *DIALEKTEKSPERT* klarer å bevise at *Sørlandsregelens* premisser er sanne for dialekt X, vil hypotesen kunne bekreftes. Først må det bevises at dialekt X er en vestnorsk dialekt. Ved å gå igjen gjennom konklusjonene i kunnskapsbasen finner systemet at *Sørlandsregelens* første premiss er identisk med *Vestlandsregelens* konklusjon. Dette fører i sin tur til at *Vestlandsregelens* premiss må bevises. Ingen annen regel har en konklusjon identisk med dette premisset. Hvis brukeren tillater systemet å be om nødvendig informasjon, vil det nå spørre om gammelnorske tostavingsord med kort rot er redusert i dialekt X (forutsatt at denne informasjonen ikke er matet inn allerede. I så fall vil det konkludere uten å spørre brukeren). Med et positivt svar på dette har *DIALEKTEKSPERT* bevist at X er en vestlandsdialekt, dvs. første premiss i *Sørlandsregelen*. Etter dette står annet premiss i *Sørlandsregelen* for tur, altså hvorvidt svake hunkjønnsord i bestemt form entall ender på *a*. Hvis det forholder seg slik testes til slutt tredje premiss i regelen. Hvis alle premissene lar seg bevise, konkluderer *DIALEKTEKSPERT* med at dialekt X er en sørlandsdialekt. I det øyeblikk et av premissene falsifiseres vil systemet automatisk slutte at dialekten ikke er en sørlandsdialekt.

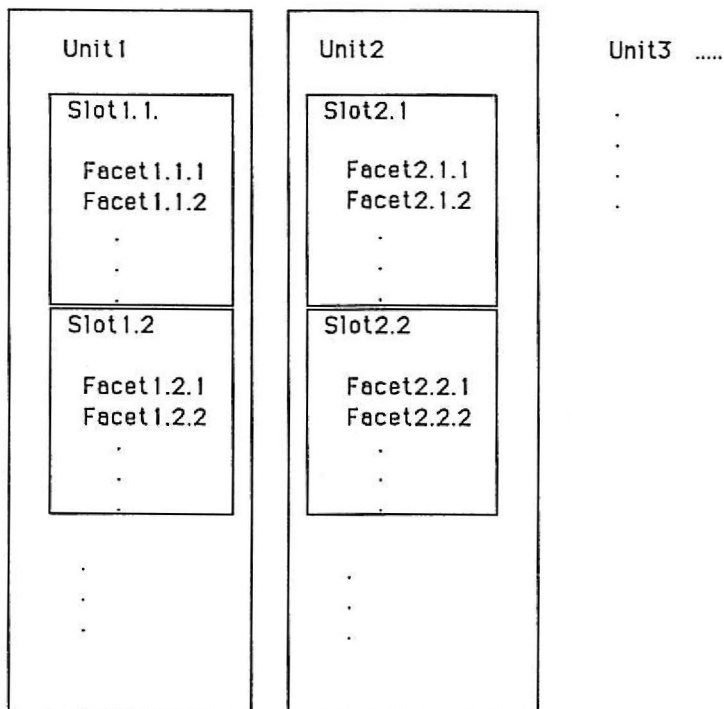
Dette eksemplet burde ha gitt et noenlunde godt inntrykk av hvordan *DIALEKTEKSPERT* fortøner seg for en tilfeldig bruker, og det skulle ha kommet klart frem at det er systemets kunnskap om norske dialekter som leder til konklusjonene.

III. LITT OM HVORDAN DIALEKTEKSPERT ER IMPLEMENTERT

DIALEKTEKSPERT er utviklet vha. ekspertsystemskallet KEE (Knowledge Engineering Environment). KEE kjøres på LISP-maskiner, dvs. datamaskiner som er spesialbygget for programmeringsspråket LISP. Ved Universitetet i Bergen er KEE tilgjengelig på en Texas Instrument EXPLORER LISP-maskin.

Kunnskap representeres i KEE som objekter som er beskrevet etter visse retningslinjer. Hvert objekt kalles en *Unit*. En unit beskrives av en eller flere *Slots*. En slot beskrives i sin tur av en *Facet*:

Fig. 3



La oss se på et konkret eksempel. I kunnskapsbasen er *Sunnhordland* en unit som bl.a. er beskrevet av slot'ene *PTK.TIL.BDG.* og *HISTORISK.A*⁽⁴⁾ Begge disse slot'ene har en facet kalt *value* som gjerne har en passende verdi⁽⁵⁾. I dette eksemplet har *value* verdien *nei* for slot'en *PKT.TIL.BDG.*, mens *value* er *a* for slot'en *HISTORISK.A*. Systemutviklerne kan legge inn så mange slot'er de måtte ønske. KEE tilbyr visse predefinerte facets som f.eks. hvilken klasse av verdier en slot-value (altså facet'en *value*) kan ta sin verdi fra (en slik klasse må naturligvis spesifiseres av systemutviklerne), hvor mange verdier en value-facet kan ha, hvordan verdien arves osv. (mer om arving av verdier nedenfor). Systemutviklerne kan lage egne facets hvis det skulle være behov for det. La oss nå «visualisere» Sunnhordlands-eksempelet:

Fig. 4

Unit Sunnhordland

Slot PTK.TIL.BDG

Inheritance: Override
Value Class: (ONE.OF JA NEI)
Max cardinality: 1
Value: NEI

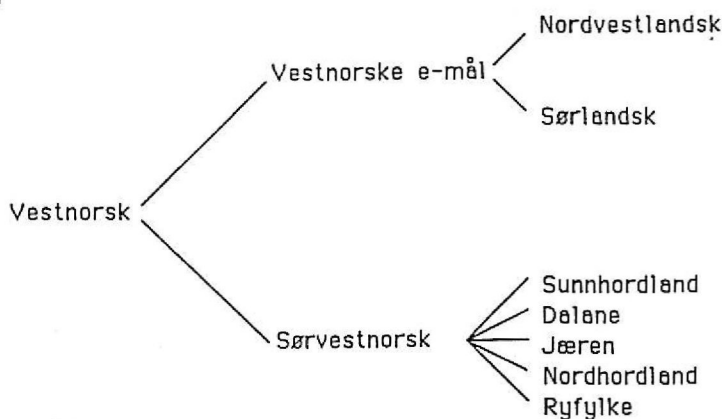
Slot HISTORISK.A

Inheritance: Override
Value Class (ONE.OF A E)
Max cardinality: 1
Value: A

At arvingen skjer ved «Override» betyr at denne slot'en kan arve verdien for den samme slot'en til «moren», dvs. den klassen *Sunnhordland* tilhører. Det er omtrent slik units kommer frem på skjermen når man jobber med KEE.

Objektene i kunnskapsbasen må naturligvis systematiseres slik at visse naturlige relasjoner mellom dem kommer frem. Det er særlig «slektskapsrelasjoner» det her er tale om. Vi ønsker f.eks. å få frem at Sunnhordlandsdialekten er en sørvestnorsk dialekt og dermed også en vestnorsk dialekt. Dette gjøres i KEE ved å operere med klasser av objekter. *Sunnhordland*, *Nord-Hordland* og *Jæren* er f.eks. subklasser av klassen *Sørvestnorsk*, og *Sørvestnorsk* er selv et objekt som er en subklasse av *Vestnorsk*. Dette kan illustreres som i fig. 5:

Fig. 5



Her er *Vestnorske e-mål* og *Sørvestnorsk* subklasser av *Vestnorsk*. *Nordvestlandsk* og *Sørlandsk* er subklasser av *Vestnorske e-mål*. *Sunnhordland*, *Dalane*, *Jæren*, *Nord-Hordland* og *Ryfylke* er subklasser av *Sørvestnorsk*. En slot kan gå i arv fra en klasse til dens subklasser eller medlemmer såfremt slot'en er deklarerert som en Member slot. Hvis man ønsker at en slot *ikke* skal arves deklarerer den som en Own slot.⁽⁶⁾ I enheten *Vestnorsk* er det en Member slot `GAMMELNORSKE.TOSTAVINGSORD.MED.KORT.ROT` som har verdien `REDUSERT`. Dermed vil denne slot'en og dens verdi dryppe ned til de enhetene den dominerer, og videre nedover hvis enhetene er subklasser av *Vestnorsk*. Denne dryppingen fortsetter så lenge der er subklasseforgreninger. Altså, `GAMMELNORSKE.TOSTAVINGSORD.MED.KORT.ROT` med verdien `REDUSERT` vil vi automatisk finne igjen som slot'er i *Vestnorske e-mål*, *Nordvestlandsk*, *Sørlandsk*, *Sørvestnorsk*, *Sunnhordland* osv. Dette er naturligvis meget nyttig når kunnskapsbasen bygges opp.

Vi har til nå sett hvordan man lager og systematiserer objekter som beskriver dialekter og dialektgrupper. Dette er ekspertsystemets statiske kunnskap om sitt domene. Reglene «setter fart» på denne kunnskapen, og vi konsentrerer oss nå om dem.

Reglene er gruppert sammen i klasser, og deres interne struktur er den samme som ved de objektene vi nettopp har sett på. Enhver regel må være medlem av en regelklasse som i sin tur er en subklasse av en spesiell predifinert regelklasse (ev. via flere subklasselinker). Dette er nødvendig for å kunne benytte ekspertsystemskallets inferenskomponent.

Vi har allerede vært borti regler og snakket om dem uformelt. Formuleringen av reglene i *DIALEKTEKSPERT* er egentlig ikke så forskjellig fra måten vi beskrev dem på i del 2. La oss ta et eksempel:⁽⁷⁾

```
(IF ((?X IS IN CLASS SOERLANDSK) AND
      (THE PTK.TIL.BDG OF ?X IS NEI))
    THEN
      (?X IS IN CLASS INDRE.SOERLANDSK))
```

En som har programmert i LISP kjenner sikkert igjen denne type syntaks, og denne regelen er faktisk formulert i LISP-dialekten Zeta-LISP. Man vil finne denne regelen som facet'en *value* i en slot *EXTERNAL.FORM* som beskriver objektet *Indre.Soerlandsregel*. Alle regelobjektene består av en rekke systemdefinerte slot'er som blir laget i det øyeblikk man velger CREATE RULE med musen fra KEE's hovedmeny. Det vil føre altfor langt å gå inn på disse slot'enes ulike funksjoner, men de må være til stede for at systemet skal kunne resonnerer vha. reglene som blir lagt inn. Premissene i reglene kan være komplekse kombinasjoner av operatorene AND og/eller OR hvis det synes påkrevet.

Hvordan reglene grupperes sammen er forsåvidt likegyldig bare de er knyttet til KEE's predefinerte regelklasser. Fordelen med å ha en fornuftig gruppering er at dette gjør editeringen og eventuelle modifiseringer av regler som behandler beslektede domener enklere. Det er enklest å editere en regelklasse der klassens regler (som er *medlemmer* av regelklassen) er listet opp. I *DIALEKTEKSPERT* finnes f.eks. regelklassen *Soerlige.e.maalregler*. Denne klassen har medlemmene *Soerlandsregel*, *Indre.soerlandsregel*, *Ytre.soerlandsregel* og *Setesdalsregel*. Hvis man nå ønsker å modifisere eller legge til noen regler som omfatter sørlandske dialekter, går man simpelthen inn og editerer *Soerlige.e.maal.regler*.

Det tør være åpenbart at rekkefølgen til en regels premisser er av stor betydning for hvor effektivt systemet bruker dem, og det er også vesentlig å unngå at systemet må teste et og samme premiss flere ganger (f.eks. ved at et premiss går igjen i mange regler på ulike nivåer i regelhierarkiet). Når det ekspertsystemet man holder på å utvikle når et visst kompleksitetsnivå, kan det være vanskelig å avgjøre hvor eventuelle flaskehalsen befinner seg. Det er her ekspertsystemskallets fordeler virkelig kommer til syne fordi man via visse tracingmekanismer faktisk kan se hvordan systemet arbeider. Man kan f.eks. prøve en backward chain og få opp et eget vindu på skjermen der man steg for steg kan se hvordan reglene virker sammen. Fig. 2 er forsåvidt

et eksempel på en slik trace (i en virkelig trace vil premisser og konklusjoner ha Zeta-LISP-syntaks, jfr. regeleksemplet ovenfor). Her kommer det for øvrig frem at første og annet premiss i *Soerlandsregelen* kanskje bør bytte plass fordi systemet slipper å kopiere *Vestlandsregelen* og sette inn variabelen *Dialekt X* hvis andre eller tredje premiss er falskt. Når det blir mange slike operasjoner i en lang resonneringsrekke kan det være lurt å foreta slike avveininger.

La oss til slutt ved å bruke et eksempel gå mer i detalj på hvordan systemet foretar en kombinert data- og hypotesedrevet resonnering. Data mates inn til *DIALEKTEKSPERT* ved LISP-kommandoer, f.eks.

```
(ASSERT '(THE HISTORISK.A OF MAALPROEVETEST IS E)
        'DIALEKTREGLER)
```

Som vi ser tilbyr KEE en kvasi-naturlig-språk syntaks (på engelsk). Kommandoen betyr at slot'en *HISTORISK.A* i enheten *MAALPROEVETEST* (altså målprøven som systemet skal finne ut av), har verdien *e*; eller mer presist, facet'en *value* i slot'en *HISTORISK.A* har verdien *e*. *DIALEKTREGLER* er regelklassen systemet skal jobbe med, og i *DIALEKTEKSPERT* dominerer denne regelklassen alle andre dialektregler. Systemet forsøker først å finne ut om det allerede har et objekt *MAALPROEVETEST*. Siden det ikke er tilfelle blir objektet *MAALPROEVETEST* laget av systemet selv. Dessuten lages en slot under *MAALPROEVETEST* som heter *HISTORISK.A*, og denne slot'ens valuefacet får verdien *e*. Så forsøker systemet å konkludere med et eller annet på bakgrunn av denne informasjonen ved å se om det er noen regler som har den inntastede informasjonen som eneste premiss. Slik forholder det seg ikke, og systemet foretar seg ikke noe mer før det får flere kommandoer. La oss nå legge inn ytterligere informasjon om målprøven:

```
(ASSERT '(THE PTK.TIL.BDG OF MAALPROEVETEST IS NEI)
        'DIALEKTREGLER)
```

Nå vil det bli laget en *PTK.TIL.BDG*-slot under objektet *MAALPROEVETEST* med verdi *nei*. Systemet kan ennå ikke konkludere med noe ut fra de opplysningene det hittil har fått, og det vil følgelig vente på mer informasjon. Anta at det blir opplyst at

```
(ASSERT '(THE SVAKT.HUNKJOENN.BESTEMT. FORM.
        ENTALL OF MAALPROEVETEST IS A)
        'DIALEKTREGLER)
```

og

(ASSERT (THE GAMMELNORSKE.TOSTAVINGSORD.MED.
KORT.ROT OF MAALPROEVETEST IS REDUSERT)
'DIALEKTREGLER)

Nå kan systemet sette i gang reglene. *Vestlandsregelen* er aksessibel siden GAMMELNORSKE.TOSTAVINGSORD.MED.KORT.ROT IS REDUSERT matcher regelens premiss. Dette fører til at kunnskapsbasen oppdateres ved at MAALPROEVETEST blir et medlem av klassen *Vestnorsk*. Hvis det er slik at objektet *Vestnorsk* har Member-slot'er, vil MAALPROEVETEST arve disse. På denne måten vil *DIALEKTEKSPERT* selv tilføre MAALPROEVETEST nye opplysninger.

Anta nå at brukeren har mistanke om at MAALPROEVETEST er en indre sørlandsdialekt. Da spør hun simpelthen om det:

(QUERY '(MAALPROEVETEST IS IN CLASS
INDRE.SOERLANDSK) 1
'DIALEKTREGLER T)

Parameteret *1* betyr at brukeren ønsker ett svar. *T* betyr at et flagg ASKUSERFLG settes til sant, og systemet kan med dette stille spørsmål hvis det ikke har nok informasjon til å avgjøre om hypotesen kan verifiseres eller falsifiseres.⁽⁸⁾ Hypotesen er identisk med konklusjonen til *Indre.soerlandsregel*. Denne regelens premisser er at ?X IS IN CLASS SOERLANDSK og THE PTK.TIL.BDG OF ?X IS NEI. Første premiss bevises ved å anvende *Soerlandsregelen* (jfr. fig. 2). Følgelig må *Soerlandsregelens* premisser bevises. Premiss 1 og 2 er allerede kjent mens det tredje er ukjent. Brukeren blir derfor spurt om det forholder seg slik at THE PALATALISERING OF MAALPROEVETEST IS NEI. Hvis brukeren svarer bekreftende er *Soerlandsregelens* konklusjon sann, og dermed er også første premiss i *Indre.soerlandsregel* bevist. Premiss 2 i sistnevnte regel er allerede kjent som sant, og *DIALEKTEKSPERT* svarer brukeren at MAALPROEVETEST er en indre sørlandsdialekt. Kunnskapsbasen er også oppdatert slik at MAALPROEVETEST er medlem av klassen *Soerlandsk* og *Indre.Soerlandsk*. MAALPROEVETEST arver dermed alle Member-slots i disse to klassene.

IV. OPPSUMMERING OG NOEN KONKLUDERENDE BETRAKTNINGER

Vi har vist hvordan *DIALEKTEKSPERT* benytter kunnskap om norske dialekter til å klassifisere målprøver. Denne kunnskapen er formulert som regler etter bestemte retningslinjer. Det er vist hvordan reglene virker sammen og hvordan kunnskapsbasen oppdateres etter hvert som nye målmerker legges inn. Vi har også sett hvordan systemet selv tilfører nye opplysninger til det objektet det skal klassifisere. Dette blir gjort ved arving av egenskaper.

Det er bare gitt en liten smakebit her av hvordan *DIALEKTEKSPERT* virker, hvordan det er bygget opp og hvilke kommandoer som tilbys av KEE. Det knyttes f.eks. gjerne kommandoer til reglene etter at de har konkludert med et eller annet. Slik kommandoer kan f.eks. være enkle beskjeder til brukeren, oppstarting av spesielle interaktive LISP-programmer laget av ekspertsystemutviklerne, matematiske utregninger osv. Slike rutiner skrives i LISP, og alt man kan uttrykke i LISP kan man uttrykke her og samtidig holde seg innenfor rammene av ekspertsystemskallet. Dette fører selvsagt til at verktøyet er meget fleksibelt.

Noen kritiske og advarende bemerkninger er på sin plass. Ekspertsystemteknologien er på ingen måte passende til et hvilket som helst komplisert problem. Før ekspertsystemteknologi tas i bruk må det være noenlunde klart at kunnskapen det skal jobbes med lar seg redusere til regler. Dette vil på ingen måte være tilfelle for flere humanistiske disipliner, særlig der fortolkning eller opplevelse er fagets egentlige anliggende.

Dessuten synes det å være en utbredt oppfatning blant de som jobber med KEE her i Bergen at programmene går temmelig langsomt og at det er lett å miste styringen når systemet begynner å bli komplekst. Dette siste skyldes antagelig at det er mange rutiner som utløses «i kulissene» av KEE uten at ekspertsystemutviklerne er oppmerksomme på dem, men det er vel grunn til å tro at slike problemer reduseres etter hvert som erfaringen med skallet blir større.

Produsenter av ekspertsystemskall reklamerer gjerne med at det ikke er nødvendig å kunne programmeringsspråket som skallet er bygget på for å kunne lage ekspertsystemer vha. det. Dette har jeg problemer med å akseptere. For meg synes det helt klart at ekspertsystemutviklere som vil benytte KEE må kunne LISP selv eller ha tilgang til en LISP-kyndig person når problemer oppstår.⁽⁹⁾

Helt til slutt, hvilken verdi har egentlig *DIALEKTEKSPERT* for

humanistisk forskning? Jeg mener at *DIALEKTEKSPERT* er et eksempel på at humanistiske innsikter lar seg implementere vha. relativt avanserte regelbaserte metoder, og at det er mulig å bruke disse innsiktene i et praktisk innrettet datasystem. Fagområder der diagnostisering og/eller systematisering står sentralt, er potensielle bruksområder for ekspertsystemteknologi; eksempelvis når det jobbes med visse typer historisk kildemateriale.

At brukeren kan følge et ekspertsystems resonneringsadferd tror jeg vil vise seg nyttig i pedagogisk sammenheng i fremtiden. Eleven/studenten eksperimenterer med systemet, og ved å følge resonneringsmønstrene lærer hun det systemet kan. Jeg er f.eks. overbevist om at det er mulig å lære ganske mye om dialektografi ved å eksperimenterer med *DIALEKTEKSPERT*. Det er således grunn til å tro at ekspertsystemer kan være relevante for datastøttet læring.

NOTER

Takk til Sperry som har donert LISP-maskinen EXPLORER og ekspertsystemskallet KEE til Universitetet i Bergen. Jeg skylder også Øyvind Ytrehus en takk for gode råd og vink angående bruk av EXPLORER og KEE. Jeg har hatt gleden av å presentere *DIALEKTEKSPERT* på et lingvistikkseminar arrangert av Institutt for fonetikk og lingvistikk ved Universitetet i Bergen. Systemet er også presentert ved MONS II i Trondheim, november 1987. Takk til publikum begge steder for kommentarer og kritikk.

1. Sitatet er hentet fra Richard Forsyth (1984) som er en instruktiv og etter min mening god innføring i og presentasjon av ulike ekspertsystemer.
2. Dermed forutsettes det altså at brukeren vet hva målmerker er. Det er en fordel, men ikke en forutsetning, at leseren har et minimum av innsikt i dialektografi for at hun skal få et godt inntrykk av *DIALEKTEKSPERT* i det som følger. Det er tilstrekkelig om man vet hva palatalisering, jamvekt, «ptk til bdg» o.l. er.
3. Jeg har selv skrevet reglene uten annen «eksperthjelp» enn Helleland & Papazian (1973). Siden jeg ikke er noen dialektekspert er det meget sannsynlig at reglene ville ha vært bedre hvis kunnskapsbasen hadde vært utarbeidet sammen med en ekspert.
4. Notasjonen PTK.TIL.BDG og HISTORISK.A er nødvendig av LISP-interne grunner. De må skrives som en sammenhengende streng for at LISP skal tolke dem slik vi ønsker.
5. *Value* kan også ha verdien «Unknown».
6. Linkingen mellom objektene i kunnskapsbasen er av to slag, enten en subklasselink (som i våre eksempler) eller en «member-link», altså at et objekt er medlem eller element av en klasse. Jfr. sett-teori der det skilles mellom elementer i en mengde og mengder av en mengde. I KEE vil alle Member slot'er bli omgjort til Own slot'er over en «Member link». Derfor vil det være slik at arving fra et objekt til et «fjernt» objekt må skje over subklasselinker.
7. Når jeg heretter refererer til regler vil jeg gjengi dem slik de faktisk er formulert i *DIALEKTEKSPERT*. Dette er grunnen til at æ, ø og å noteres som ae, oe og aa i reglene.

8. Det er mulig å bruke variabler i slike hypoteser, f.eks.

```
(QUERY '(MAALPROEVETEST IS IN CLASS ?X) I
        'DIALEKTREGLER NIL)
```

Siden ASKUSERFLG her er NIL kan systemet ikke be om tilleggsopplysninger. Antall svar bestemmes av bruker, f.eks. 1 eller 'ALL. Hvis 'ALL velges vil alle klasser MAALPROEVETEST tilhører bli listet opp. Hvilke svar man får i en slik hypotese med variabler og systemet bare skal gi ett svar, er avhengig av om systemet jobber med reglene i en «depth-first» eller «breadth-first» modus.

9. Da jeg var på kurs i England hos Sperry for å lære KEE ble det *forutsatt* at kursdeltagerne var fortrolige med LISP.

REFERANSER

Helleland, B. & E. Papazian, 1973. «Oss sei oss, oss», NRK — Skoleradioen 14 — 73, Oslo.

Forsyth, R. (red.), 1984. *Expert Systems. Principles and Case Studies*, Chapman and Hall, London.

Torbjørn Nordgård er universitetsstipendiat i datalingvistikk ved Inst. for fonetikk og lingvistikk, Universitetet i Bergen.

AKTUELLE METODER FOR AUTOMATISK MORFOLOGISK ANALYSE OG MULIGE ANVENDELSER AV DEM I FREMTIDENS TEKSTBEHANDLINGSSYSTEMER

Tove Fjeldvig

Følgende tekst ble gitt som prøveforelesning til magistergradseksamen i informasjonsvitenskap, Bergen 1. juni 1987.

1. INNLEDNING

Formålet med en morfologisk analyse er å finne fram til de enkelte morfemene i et ord. Ordet *GUTTEN* er for eksempel bygd opp av to morfemer: *GUTT* og *EN* — hvor *EN* er en bøyningssendelse i bestemt form entall.

Ethvert ord på norsk vil ha en morfemstruktur som passer inn i følgende formel:

(1) $pre^* \text{ rot } avl^* ((E/S) \text{ pre}^* \text{ rot } avl^*)^* \text{ b\ae}yn^*$

hvor

pre^* : 0, 1 eller flere prefikser
 rot : rot
 avl^* : 0, 1 eller flere avledningsendelser
 $b\ae}yn^*$: 0, 1 eller flere bøyningssendelser
(E/S) : mulige fuge-E eller -S
 $((E/S) \text{ pre}^* \text{ rot } avl^*)^*$: 0, 1 eller flere forekomster
av $((E/S) \text{ pre}^* \text{ rot } avl^*)^*$

Eksempel

(1) **GUTT -EN**
 ↑ ↑
 rot bøyn

(2) **TEKST BE- HANDL -ING -EN**
 ↑ ↑ ↑ ↑ ↑
 rot pre rot avl bøyn

Denne formelen vil kunne danne utgangspunktet for en automatisk morfologisk analyse av norske ord, og tilsvarende uttrykk vil også kunne settes opp for andre språk.

Det oppstår en rekke ulike problemer når man skal forsøke å få en datamaskin til å identifisere de ulike morfemene i et ord. Eksempel på denne type problemer oppstår når:

- (1) Ulike morfemer er representert ved morfer som har overlappende bokstavstrenger, f.eks. *EN* er en bøyningssendelse i ordet *GUTTEN* og en del av roten i ordet *VERDEN*.
- (2) Like morfer representerer ulike morfemer, f.eks. *ER* er en presensendelse i verbet *LESER* og en ubestemt flertallsendelse i substantivet *GUTTER*.
- (3) Enkelte ord forandrer stamme ved bøyning (fonologiske og morfologiske vekslinger), f.eks. *REGEL* — *REGLER* (fonologisk veksling) og *FAR* — *FEDRE* (morfologisk veksling).
- (4) Enkelte sammensatte ord har en fugebokstav, mens andre har det ikke, f.eks. *BILAVGIFT* har ingen fugebokstav, mens *EKSAMENSAVGIFT* har fuge-S og *BARNEBIDRAG* har fuge-E.

Omfanget av problemene vil variere fra språk til språk. Fugeproblemet i punkt (4) er f.eks. et stort problem på norsk, men ikke på engelsk. Likeledes er fonologiske vekslinger et langt større problem på finsk enn på norsk og engelsk.

De fleste metodene for automatisk morfologisk analyse er utviklet med utgangspunkt i en konkret anvendelse, f.eks. informasjonssøking, vokabularstudier, ordbokproduksjon, orddeling eller syntaksanalyse. De er derfor ofte rettet mot bestemte typer morfemer. I f.eks. informasjonssøking er man interessert i å få likestilt ord med felles rot, mens man i en syntaksanalyse har behov for kunnskap om hvilken ordklasse et ord tilhører.

Ved vurdering av denne type metoder må man derfor ta utgangspunkt i hva som er formålet med metoden. I f.eks. informasjonssøking vil det kunne være forutsetning at metoden ikke er basert på et stammeleksikon. Et leksikon er ressurskrevende å vedlikeholde, og det kan diskuteres hvorvidt det er økonomisk forsvarlig å basere metoden på et leksikon framfor en regelbasert metode — sett i forhold til søkekvaliteten. I andre sammenhenger derimot, som f.eks. i maskinoversetting, er det en forutsetning at metoden gir lite feil.

Enhver metode for automatisk morfologisk analyse må være basert på noe informasjon om morfemene. Enkelte metoder forutsetter en oversikt over bøynings- og avledningsendelser, mens andre er basert på en fullstendig oversikt over alle morfer og deres morfosyntaktiske egenskaper. I enkelte miljøer har de utelukkende vært opptatt av metoder som ikke gjør bruk av et rot- eller stammeleksikon. Målsetting har bl.a. vært å undersøke hvor langt man kan nå ved bare å ta utgangspunkt i de ytre trekkene ved morfemene. Et eksempel på et slikt miljø finner vi i Stockholm rundt *Benny Brodda* og *Gunnel Källgren*. Begge har spilt en sentral rolle i arbeidet med automatisk morfologisk analyse i Skandinavia, og Brodda har bl.a. utviklet et eget programverktøy for automatisk morfologisk analyse, som er kalt Beta (jfr. Brodda 1979).

I andre miljøer har de hatt et mer teoretisk utgangspunkt hvor målet har vært å utvikle generelle, språkuavhengige modeller som kan beskrive morfemstrukturen i et hvilket som helst ord innenfor et gitt språk. Et eksempel på et slikt miljø finner vi i Finland med *Kimmo Koskenniemi* i spissen. Finsk er et sterkt syntetisk språk med en svært kompleks morfologi — noe som gjør en automatisk morfologisk analyse vanskelig, men også utfordrende.

I det følgende skal jeg gi eksempler på ulike metoder for automatisk morfologisk analyse. Jeg vil her skille mellom metoder som gjør bruk av et rot- eller stammeleksikon, og de som ikke gjør det.

2. AKTUELLE METODER FOR AUTOMATISK MORFOLOGISK ANALYSE

(a) Metoder som ikke er basert på et rot- eller stammeleksikon

«Suffix-cutoff» eller «suffix-stripping»

Den første tilnærming til automatisk morfologisk analyse fikk vi alt tidlig på 1960-tallet hvor det ble utviklet enkle metoder for informasjonssøking som kan karakteriseres som «stripping off the endings» (jfr. Salton 1968).

Metodene gikk i korthet ut på å kutte endelsen(e) til ordene. Det man sto igjen med, antok man var stammen til ordet, og ord med lik «stamme» ble så gruppert og behandlet under ett i søkeprosessen.

Denne metoden var basert på en liste over aktuelle endelser og endelseskombinasjoner. I enkelte tilfeller ble det også gjort bruk av noen få regler som tok hånd om fonologiske vekslinger, f.eks. at *Y* i slutten av et (engelsk) ord blir forandret til *I(E)* når dette blir etterfulgt av en endelse.

Slike metoder finner vi også eksempel på i dagens informasjonssøke-system, f.eks. i det amerikanske rettslige informasjonssystemet LEXIS. De egner seg forholdsvis bra for språk hvor stammevariasjon ved bøyning forekommer forholdsvis sjelden (f.eks. engelsk). Derimot ville de f.eks. ikke ha fungert for finsk da de ikke tar hensyn til hvilke morfemer som kan kombineres, og hvilke vekslinger som finner sted når morfemene kombineres (jfr. Koskenniemi 1983). Dessuten er disse metodene basert på en fullstendig liste over alle endelseskombinasjoner, noe som er praktisk umulig å sette opp for et språk som finsk hvor det kan være opptil 20.000 ulike bøyingsformer av ett og samme substantiv.

Fjeldvig/Goldens metode for automatisk rotlemmatisering

Denne metoden er forsåvidt basert på det samme prinsippet som «suffix-cutoff»-metoden ved at den behandler ordet bakfra og undersøker om ordets slutt er lik en endelse. Før imidlertid endelsen kuttes, vil den sjekke om det som står foran endelsen, er en rot. Hvis også dette synes å være en endelse, behandles ordet på nytt. På denne måten blir endelseskombinasjoner håndtert.

Problemet med fonologiske vekslinger er her løst ved omskrivingsregler som erstatter en tegnstreng med en annen. F.eks. *EL* i ordene

REGEL og *REGELN* blir erstattet med *L*, mens ordene *REGLER* og *REGLENE* får bare fjernet endelsen. Følgelig får de alle oppslagsformen «*REGL*».

Morfologiske vekslinger er løst ved at alle uregelmessige verb og substantiv er kartlagt, og at det er konstruert lignende omskrivingsregler for disse som for de fonologiske vekslingene.

Metoden har vært testet på ulike materialer, og resultatene har vist at den er i stand til å gruppere ord med samme rot og samme betydning i ca. 97% av tilfellene.

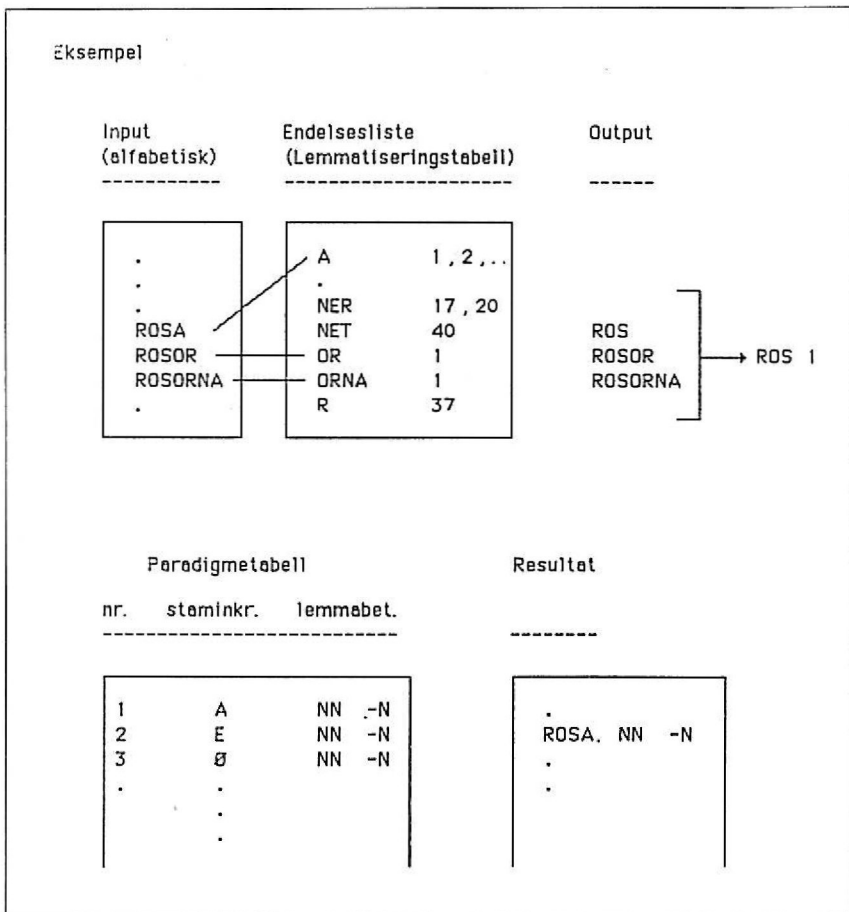
Hellbergs metode for automatisk lemmatisering

Steffan Hellberg var den første i Skandinavia som utviklet en metode for automatisk morfologisk analyse. Formålet med metoden var å lette arbeidet med utgivelse av den svenske frekvensordboka, som var basert på et stort tekstmateriale, bl.a. avisartikler. Her skulle hvert lemma representeres ved en oppslagsform og en lemmabetegning (paradigmeangivelse).

Hellbergs metode er basert på at ordene som skal analyseres, er sortert i alfabetisk rekkefølge. Den tar utgangspunkt i en liste med endelser hvor hver endelse er representert med ett eller flere paradigmenummer som angir hvilke(t) paradigme(r) ordet kan være en del av. Et eksempel på anvendelse av metoden er vist nedenfor.

Metoden sammenligner ord som følger etter hverandre i «input»-lista for å se om de innledes med samme bokstavstreng. Ved tilslag, undersøker den om «resten» av ordene er endelser ifølge endelseslista. I eksempelet har ordene *ROSA*, *ROSOR* og *ROSORNA* strengen «*ROS*» felles, og *A*, *OR* og *ORNA* er alle endelser ifølge lista. Den sjekker så om disse endelsene har et felles paradigmenummer. I dette eksempelet har alle paradigmenummeret 1, og de blir derfor ansett for å tilhøre samme lemma, og de får paradigmenummeret 1. Dette nummeret blir brukt til å slå opp i paradigmatabelen. Her finner man at paradigme 1 har betegnelsen «*NN -N*» og at oppslagsformen til lemmaet konstrueres ved at man til «*ROS*» legger til bokstav *A* — dvs. *ROSA*.

Forsøk med denne metoden har vist at 95% av alle de ulike ordene blir riktig gruppert, og at 85% av lemmaene får riktig lemmabetegning. Feil gruppering skyldtes i de fleste tilfellene fonologiske og morfologiske vekslinger, noe som det ikke var tatt hensyn til i denne metoden. Ved vurdering av resultatet bør man dessuten ta i betraktning at materialet var homografseparert, dvs. at alle homografer var kartlagt manuelt før arbeidet startet.



Også slike metoder fungerer relativt bra i språk hvor det er lite stammevariasjon, og lignende metoder har også vært utført på fransk og tysk (jfr. Koskenniemi 1983:13).

Källgrens metode for automatisk splitting av sammensatte ord

Källgren er blant de få som har utviklet en metode for automatisk splitting av sammensatte (svenske) ord. En slik metode kan være et nyttig hjelpemiddel i flere sammenhenger, f.eks. informasjonssøking, automatisk produksjon av «bak-i-boken»-register, orddeling eller automatisk maskinoversetting.

Fastsettelse av delingspunktet (evt. delingspunktene) krever kjennskap til morfemstrukturen i ordet — og en slik metode vil derfor forutsette en morfologisk analyse. Källgrens fremgangsmåte har mye til felles med Fjeldvig/Golden som er omtalt nedenfor.

For at en slik metode skal kunne brukes på en løpende tekst, må den både kunne gjenkjenne om et ord er sammensatt og splitte et sammensatt ord i de enkelte leddene. Metoden kan kort beskrives i følgende fire punkter:

- (1) Leksikalsk merking av ord av lukkede ordklasser
- (2) Morfologisk analyse
- (3) Konsonant- og vokalanalyse
- (4) Sannsynlighetsberegning pga. frekvensstudier for ulike ordstart og ordslutt

Den første fasen går ut på å merke alle ord av lukkede ordklasser som ikke er sammensatt. Prosessen tar utgangspunkt i en kort forhåndsdefinert ordliste som normalt dekker ca. 40% av de løpende ordene i en tekst. Formålet med denne prosessen er utelukkende å redusere kjøretiden.

Den morfologiske analysen i pkt. (2) merker alle affikser og bøyningssendelser i ordet — bortsett fra eventuelle prefikser i midten av ordet. Dette skjer på bakgrunn av en oversikt over alle bøyningssendelser og affikser — samt informasjon om hvilke avledningssendelser som tar fuge-S og hvilke som ikke gjør det.

Den tredje fasen er den mest omfattende. Den undersøker om de stammene som man har kommet fram til i pkt (2), må eller bør deles — og i så fall hvor. Fasen innledes med en konsonantanalyse som kan avgjøre om et ord må deles, og hvor eventuelt delingspunktet går, f.eks. ordet *TEKSTBEHANDLING* blir delt i leddene *TEKST* og *BEHANDLING* fordi konsonantkombinasjonen *STB* ikke kan forekomme i et usammensatt ord og *TB* ikke kan innlede et ord. Deretter gjennomføres en vokal- og stavelsesanalyse som signaliserer om et ord er sammensatt og hvor omtrent delingspunktet går. Dette skjer bl.a. ut fra en hypotese om hvor mange vokaler et ord kan inneholde, og hvilken rekkefølge vokalene kan ha i et flerstavet ord. Vokalene vektlegges på en slik måte at når summen av vokalvektene (regnet fra venstre mot høyre) når en viss størrelse, antar man at leddgrensen er nådd.

Den siste fasen trer i kraft hvis det er flere mulige delingspunkter. Den beregner hvilket delingspunkt som er mest sannsynlig ut fra opplysninger om frekvensen til ulike ordstart og ordslutt.

Det hevdes at metoden fungerer bra, men foreløpig er det ikke publisert resultater som viser hvor bra den er. Det er også utviklet metoder for automatisk splitting av sammensatte ord på tysk (jfr. Niedermair/Thirmair/Buttel 1984) og engelsk (jfr. Jones/Bell), men fordi disse er utviklet i kommersiell regi, foreligger det ingen detaljert beskrivelse publisert.

Fjeldvig/Goldens metode for automatisk splitting av sammensatte ord

Felles for Källgren og Fjeldvig/Goldens metode er at begge starter med en leksikalsk merking og en morfologisk analyse. Dessuten gjør begge bruk av en konsonant- og vokalanalyse, men i Fjeldvig/Goldens metode er delingspunktet ikke beregnet ut fra vokalrekkefølgen i ordet, slik som Källgren gjør. Derimot er det i Fjeldvig/Goldens metode lagt inn mer informasjon om den enkelte morf og hvilke morfer som kan kombineres. Sannsynlighetsberegningen til slutt er dels basert på vektlegging av morfene ut fra deres gjenkjennelighet, dels på antall ledd og dels på antall stavelser i hvert ledd.

Undersøkelser har vist at man med denne metoden vil finne riktig delingspunkt i ca. 87-89% av tilfellene.

Fjeldvig/Goldens metode er omtalt i Fjeldvig 1987 og Fjeldvig/Golden 1987.

(b) Metoder basert på et rot- eller stammeleksikon

Det er en grense for hvor gode resultater man kan oppnå med en metode som bare gjør bruk av morfemenes ytre struktur. Uten bruk av et rot- eller stammeleksikon vil man f.eks. ikke kunne fastslå om et ord er bøydd når bøyningsmorfemet er realisert ved en morf som har en bokstavstreng som også kan være en del av en rot (f.eks. *EN* i *VERDEN*).

Dermed er det ikke sagt at man er garantert et 100% riktig resultat ved bruk av et leksikon. Dette beror på hvilken informasjon man har lagt inn i leksikonet, og hvordan man utnytter denne informasjonen.

I det følgende skal jeg gi eksempel på to ulike metoder for automatisk morfologisk analyse som begge gjør bruk av et leksikon på hver sin måte. Formålet med metodene er å kartlegge alle morfemene i ordet. Den siste metoden kan dessuten generere de ulike bøyningsmorfemene til et ord.

SMORF er en kraftig analysator av svenske ordformer. Den er basert på et leksikon hvor alle ulike ord er representert ved en teknisk stamme og et paradigmenummer. En teknisk stamme defineres som den delen av ordet som er felles for alle ord innenfor et lemma, f.eks. er *ROS* den tekniske stammen for *ROSA(S)*, *ROSOR(S)* og *ROSOR-NA(S)*.

Paradigmenummeret henviser til en beskrivelse av paradigmet. Her finner man pekere til subrutiner som beskriver hvilke bøyingsendelser som gjelder for dette paradigmet. På den måten kan man «kjede seg fram» til riktig tolkning av ordet.

I paradigmebeskrivelsen ligger det også en test på om ordet er sammensatt. I så fall bli man henvist tilbake til leksikonet for å søke etter neste ord. Den samme prosedyren gjelder også om ordet er avledet ettersom avledningsendelsene er registrert i leksikonet på samme måte som stammer.

Eksempel på bruk av SMORF er gitt nedenfor. Systemet gir en fullstendig beskrivelse av alle morfemene i ordet. Hvis ordet har ulike tolkninger, får man en fullstendig beskrivelse av hver enkelt tolkning.

Eksempel

```
SMORF Window
96-SMORF: AVLASTA
(AVLASTA (("AV" "LASTA" (verb inf))
          ("AV" "LASTA" (verb imp))))
97-SMORF: AVSE
(AVSE (("AVSE" (verb inf))
        ("AVSE" (verb imp))
        ("AV" "SE" (verb inf))
        ("AV" "SE" (verb imp))))
98-SMORF:
```

```
SMORF Window
57-SMORF: SMÖRGÅS
(SMÖRGÅS (("SMÖRGÅS" (noun non-neut indef sing base/gen))
          ("SMÖR" "GÅS" (noun non-neut indef sing base/gen))
          ("SMÖR" "GÅ" (verb inf päs))
          ("SMÖR" "GÅ" (verb pres päs))
          ("SMÖR" "GÅ" (verb imp päs))))
58-SMORF:
```

En svakhet ved SMORF er at det ofte gir en overproduksjon av usannsynlige løsninger. Dette har dels sammenheng med at den skal kunne tolke et hvilket som helst ord uansett om det f.eks. er en forkortelse eller et ikke-svensk ord, og dels at det ikke har innebygget morfosyntaktiske regler.

Koskenniemi «2-nivå»-modell

Det har tidligere vært nevnt at finsk har en svært kompleks morfologi. I det følgende skal jeg kort presentere en metode som også kan analysere finske ord. Metoden, som har vakt stor interesse også utenfor Finlands grenser, er utviklet av Kimmo Koskenniemi, og den har fått tilnavnet «2-nivå»-modellen (jfr. Koskenniemi 1983).

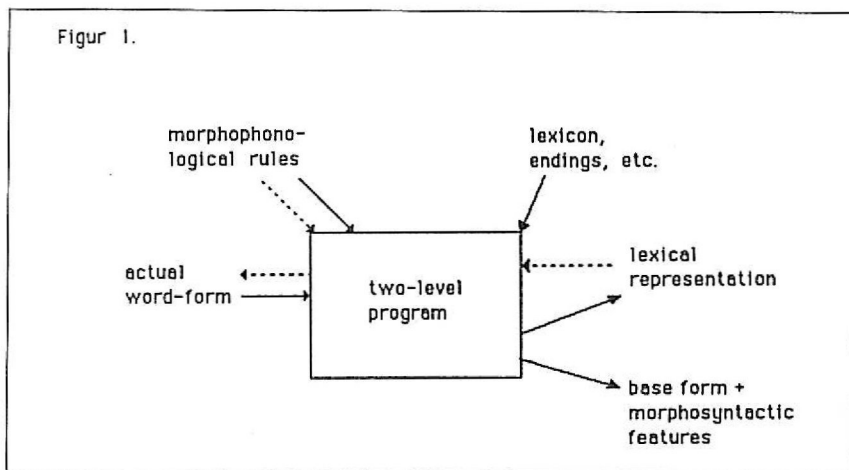
Metoden kan man se på både som (1) en teori og en formalisme for beskrivelse av orddannelse (f.eks. bøyninger, avledninger og sammensetninger) og (2) et datamaskinprogram for ordanalyse og syntese. Modellen er språkuavhengig, og den kan både brukes til å gjenkjenne alle morfemene i et ord og til å produsere de ulike bøynings- og avledningsformene til et gitt ord.

En grov skisse av «2-nivå»-modellen er gitt nedenfor (jfr. Koskenniemi 1985:2). De heltrukne pilene illustrerer at man kan gi en hvilken som helst ordform som inndata og få ut grunnformen til ordet samt beskrivelse av de morfosyntaktiske egenskapene til ordet. Til dette bruker man et sett med morfologiske regler og et leksikonsystem. De stiplede pilene viser at man også kan gå den motsatte veien, dvs. ta utgangspunkt i grunnformen til et ord samt en beskrivelse av de morfosyntaktiske egenskapene, og få ut den aktuelle ordformen.

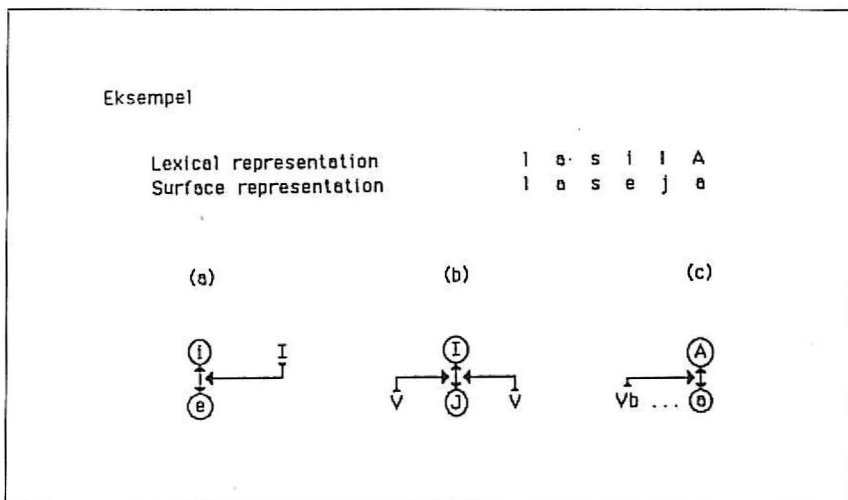
Formalismen gir muligheten til å beskrive ordstrukturen i et gitt språk. Den består av to komponenter: et leksikonsystem og en samling med regler. Leksikonet er delt opp i flere subleksikon. Ett av dem inneholder (ord)røtter, mens de øvrige er reservert for affikser og bøyningsendelser. Mulige rekkefølger av disse elementene er tatt vare på gjennom et lenkesystem. Det er imidlertid viktig å merke seg at denne rekkefølgen bare angir en «leksikalsk representasjon» av de enkelte ordformer som kan være forskjellig fra den faktiske formen til et ord (f.eks. *REGEL* + *ER* blir til *REGLER*).

Reglene definerer hvordan denne «leksikalske representasjonen» kan realiseres til ordformer. De refererer til denne representasjonen av ordet og angir hvilke betingelser som må være oppfylt for at en fonologisk veksling skal finne sted og hvilke vekslinger som i så fall finner sted. Man snakker her om parallelle regler fordi de både opererer på

Figur 1.



den «leksikalske» og den faktiske representasjonen av ordformen samtidig. Fremgangsmåten lar seg best illustrere ved et eksempel som er hentet fra Koskeniemi 1985:8:



Det finske ordet *laseja* blir brukt som eksempel (jfr. «surface representation» nedenfor). I rotleksikonet er dette ordet representert ved «lasilA» hvor I og A er symboler som har en spesiell betydning, og som ikke eksisterer i de faktiske ordformene. Symbolet I er brukt for å symbolisere at flertalls morfemet vil kunne ha en spesiell effekt på

den vokalen som står foran. I eksempelet ovenfor ser vi av regel (a) at vokalen *i* går over til *e* når det som følger bak er et flertalls morfem.

Normalt vil flertalls morfemet for dette ordet realiseres som *i*, men hvis det står en vokal (V) på hver side av *i*'en, vil den gå over til *j* (jfr. regel (b) i eksempelet).

Symbolet A i den leksikalske representasjonen står for et arkifonem som kan realiseres som enten *a* eller *ä* avhengig av kvaliteten på vokalen i stammen foran. Hvis stammen har en bakre harmoniserende vokal, vil A gå over til *a* (jfr. regel (c)).

På denne måten beskriver «2-nivå»-modellen morfologiske variasjoner.

3. MULIG ANVENDELSE AV MORFOLOGISK ANALYSE I FRAMTIDENS TEKSTBEHANDLINGSSYSTEMER

Det finnes mange anvendelsesmuligheter for automatisk morfologisk analyse i et tekstbehandlingssystem. De mest nærliggende er orddeling og stavekontroll, men jeg ser også nytten av denne type metoder i forbindelse med:

- automatisk stikkordproduksjon («bak-i-boken»-register)
- avsløring av inkonsistent språkbruk, uheldige stilistiske uttrykksmåter og syntaksfeil
- grensesnitt i naturlig språk
- tale som inndata (diktering)

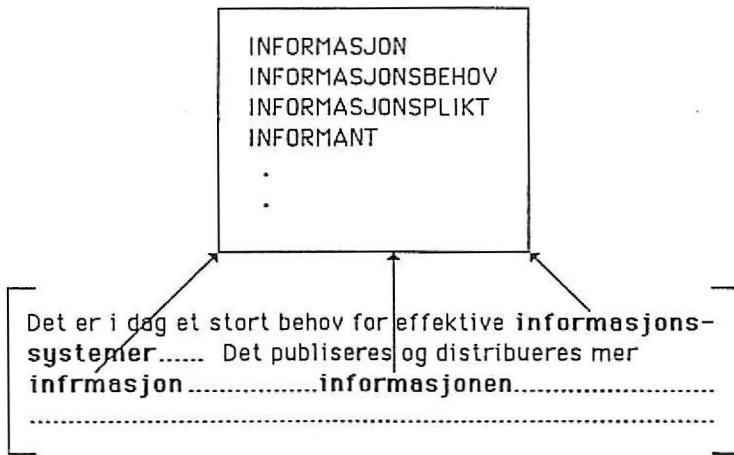
Både stavekontroll, orddeling og automatisk stikkordproduksjon finnes i flere av dagens systemer, men bortsett fra noen få eksempler på automatisk orddeling, er det ingen av disse som gjør bruk av morfologisk analyse.

Stavekontroll («Spelling»)

I dagens systemer foregår stavekontroll på den måten at alle ord sammenlignes mot en forhåndsdefinert ordbok. I f.eks. tekstbehandlingssystemet WordPerfect er det lagt inn en ordbok på ca. 60.000 ulike norske graford. Fremgangsmåten kan variere, men det mest vanlige er at systemet stopper opp ved ord som ikke forekommer i ordboka (f.eks. *INFRMASJON*, *INFORMASJONEN* og *INFORMASJONSSYSTEMER* i figuren nedenfor), og spør om brukeren aksepterer dette ordet eller ønsker å korrigere det. Hvis ordet aksepteres, får brukeren

også muligheten til å legge det til ordboka. Hvis ordet derimot skal rettes opp, får brukeren fram en oversikt over de ordene i ordlista som «ligner» på dette ordet. Ved hjelp av piltastene kan man peke på det riktige ordet hvis dette finnes, og dette vil automatisk bli erstattet med det feilskrevne ordet. I eksempelet nedenfor vil ordet *INFRMASJON* bli erstattet med *INFORMASJON*.

Eksempel



Denne fremgangsmåten kan gjøres bedre på mange måter ved bruk av automatisk morfologisk analyse. Med kjennskap til hvilke konsonantsekvenser som kan innlede, avslutte eller opptre i midten av en morf, vil man kunne sjekke om ordet har en «normal» oppbygning, f.eks. *INFRMASJON* har det ikke. En slik fremgangsmåte vil imidlertid bare kunne avsløre en del av stavefeilene — og vil derfor bare være en «grovkorrektur» før en eventuelt sjekk mot ordboka.

Slik stavekontroll fungerer i dag, vil man aldri kunne få en ordbok som dekker alle ord. Systemet oppfatter to ord som forskjellige hvis minst én av bokstavene eller rekkefølgen av dem er forskjellig. Det vil f.eks. stoppe opp på ordet *INFORMASJONEN* selv om ordet *INFORMASJON* ligger i ordboka. Det samme skjer ved bruk av et sammensatt ord (*INFORMASJONSSYSTEMER*) selv om de enkelte usammensatte ordene forekommer i ordboka.

En automatisk morfologisk analyse ville kunne gjøre oppdateringen av ordboka enklere. Hvis man f.eks. hadde erstattet ordboka med et stammeleksikon av Hellbergs modell, ville man i alle fall ha fått dekket alle bøyningsformene til stammen. Den beste løsningen ville imidlertid ha vært en metode som «2-nivå»-modellen som også dekker avledningsformer og sammensatte ord. Problemet er imidlertid at også denne metoden er basert på et leksikon som må oppdateres av og til — og dette må gjøres av personer som vet hvordan. En løsning kunne være å legge alle ord som ikke var dekket av leksikonet ut på en unntaksordliste, som ble tatt hånd om av leverandøren med jevne mellomrom.

Fjeldvig/Goldens metode for automatisk rotlematisering vil langt på vei kunne løse problemene med bøynings- og avledningsendelser. Man kan her erstatte ordboka med ei «rotbok» som oppdateres automatisk. Det vil si at ved stavekontroll reduserer hvert ord til en oppslagsform som i sin tur sjekkes mot «rotboka». Ved oppdatering blir oppslagsformen til ordet bare lagt til «rotboka». Selv om denne metoden kan gi en ukorrekt oppslagsform, vil dette sjeldent få innflytelse på stavekontrollen. Derimot vil den akseptere feilskrevne ord av typen *HUSA*, *GUTTET* etc., fordi metoden ikke er basert på informasjon om hvilket paradigme ordet tilhører.

Orddeling

De fleste orddelingsprogrammer er basert på en ren bokstavmanipulasjon. De tilpasser delingspunktet ut fra hvor mange tegn det er igjen på linja og ut fra kjennskap til informasjon om hvordan ulike konsonantklustre skal deles mellom to vokaler. Enkelte utenlandske programmer gjør bruk av morfologisk informasjon til å identifisere bøyningsendelser og affikser. I f.eks. Danmark er det utviklet denne type programmer som opererer med en feilprosent på bare 1.8% uten bruk av unntaksordbok (jfr. Hansson 1979).

For å kunne oppnå best mulig orddeling, må programmet både kunne gjenkjenne eventuelle bøyningsendelser og affikser i ordet — og splitte sammensatte ord i de enkelte ledd. Det siste er det vanskeligste fordi dette forutsetter en morfologisk analyse som kan kartlegge grensene mellom de ulike morfene i ordet.

Dette problemet kan løses med bruk av de metodene som er utviklet for automatisk splitting av sammensatte ord. Det kan for øvrig nevnes at Golden og jeg for tiden arbeider med å utvikle en metode for automatisk orddeling for norsk, som er basert på vår metode.

Automatisk stikkordproduksjon

Flere av dagens tekstbehandlingssystemer gir muligheten til å bygge opp et «bak-i-boken»-register. Dette skjer ved at brukeren merker de ordene i teksten som skal med i registeret. Systemet vil så generere et alfabetisk register hvor hvert ord er angitt med henvisninger til hvilke sider i dokumentet ordet forekommer. Et eksempel er vist nedenfor.

Eksempel

ORDDELING	s. 12, 14
ORDDELINGEN	s. 12, 13, 67
ORDDELINGSPROGRAM	s. 67, 88
DELING AV ORD	s. 12, 65

Denne prosessen vil kunne forbedres ved bruk av automatisk morfologisk analyse. For det første vil den kunne gruppere ord innenfor samme rotlemma (eller lemma). I dag må brukeren selv slå sammen de ulike bøynings- og avledningsformene i registeret. En metode for automatisk rotlemmatisering ville kunne ha løst dette problemet, men hvis man f.eks. skulle ha brukt Fjeldvig/Goldens metode, måtte man først ha modifisert reglene slik at oppslagsformen for rotlemmaet ble lik grunnformen til ordet. I dag er det 90% av ordene som får en oppslagsform som er lik grunnformen.

For det andre ville det ha vært ønskelig å fått likestilt sammensatte ord med fraser av samme betydning, f.eks. *ORDDELING* og *DELING AV ORD*. Til denne oppgaven trenger man både en morfologisk, syntaktisk og semantisk analyse. Den morfologiske analysen må kunne identifisere alle de ulike morfemene i ordet ettersom den både skal splitte et sammensatt ord og informere om hvilken ordklasse et ord tilhører. Det siste er nødvendig informasjon for den syntaktiske analyse som har til oppgave å gjenkjenne frasene, f.eks. en preposisjonsfrase som i eksempelet ovenfor. Den semantiske analyse vil gå ut på å undersøke om den aktuelle frasen har samme meningsinnhold som det

sammensatte ordet. Dette er umulig uten informasjon om den semantiske relasjonen mellom ordene, og en mer realistisk løsning ville ha vært å overlate denne avgjørelsen til brukeren.

For det tredje kunne man tenke seg et slikt register bygd opp hierarkisk, f.eks. la ordet *ORDDELINGSPROGRAM* komme inn som en undergruppe til ordet *ORDDELING*. Dette vil bl.a. forutsette et program for automatisk splitting av sammensatte ord.

Avsløring av inkonsistent språkbruk

Med inkonsistent ordbruk tenker jeg i første rekke på eksempler av typen *JENTA* — *JENTEN* og *BYGGET* — *BYGDE*. På bakgrunn av en morfologisk analyse som kan identifisere bøyingsmorfemet (eller morfemene) til et ord, vil man kunne oppdage utilsiktede variasjoner av denne typen.

Dette problemet er det samme som Hellberg har forsøkt å løse i sin metode for automatisk lemmatisering (jfr. foran). Denne metoden ga riktignok feil lemmabetegning i 15% av tilfellene, men den gjorde ikke bruk av et leksikon. Hvis man skal stille strengere krav til resultatet, så må man ta i bruk en metode som er basert på et rot- eller stammeleksikon. Eksempel på en slik metode er Koskenniemis «2-nivå»-modell.

Stilistisk uheldige uttryksmåter

Et eksempel på denne type anvendelse ville være å gi brukeren beskjed hvis det samme leksemet ble gjentatt mange ganger i f.eks. et avsnitt. Ved bruk av en metode for automatisk rotlemmatisering ville man kunne gruppere ord innenfor samme rotlemma, og føre statistikk over hvor ofte det enkelte rotlemma var representert i teksten. Hvis den samlede ordfrekvensen til et rotlemma ble for høy i forhold til det totale antall løpende ord innenfor et tekstintervall, kunne brukeren få beskjed om dette og bli oppfordret til å variere språkbruken.

Syntaksfeil (samsvarsbøyning)

I et fremtidig tekstbehandlingssystem kan man tenke seg ulike former for syntakskontroll. Et eksempel er kontroll av samsvarsbøyning mellom ord, f.eks. avsløre feil som *EN GUTTEN*, *DE RØDE BALL* eller *DET RØDE BALLEN*. Et annet eksempel er kontroll av infinitivsfraser, f.eks. avsløre feil som *Å DANSER* eller *FOR OG DANSE*.

Begge disse tilfellene forutsetter en morfologisk analyse som kan identifisere bøyningsmorfemene i ordet, og en syntaktisk analyse som kan sjekke om bøyningene er i samsvar.

4. AVSLUTNING

Rammen for denne delen av foredraget har vært fremtidens tekst-behandlingssystemer. På ett eller annet tidspunkt vil man slippe å bruke tastatur eller mus, og i stedet diktere teksten rett inn og gi alle kommandoene muntlig. Dette forutsetter at systemet både kan oppfatte de enkelte ordene og den syntaktiske sammenheng mellom dem. For å få dette til, må man ha en god metode for automatisk morfologisk analyse. Det er ikke vanskelig å si *at* denne type produkter kommer — snarere *når* det kommer. I henhold til *Sam Fuller*, som er forskningsdirektør for ett av de største dataselskapene i verden og disponerer 1 mill. dollar årlig, så vil det ta minst 15 år til før man kan lansere produkter som dette. Tiden får vise om han har rett.

LITTERATUR

- Andersson Östling, Annette (1986) «A two-level description written French». Karlsson s. 195-202.
- Brodde, Benny (1979) «BETA-systemet som verktøy for lingvistiske experiment — morfologisk analys enligt ytkriterier». Mægaard (1979) s. 103-122.
- Fjeldvig, Tove/Golden, Anne (1984) *Automatisk rotlematisering — et lingvistisk hjelpemiddel for tekstsøking*. Institutt for rettsinformatikk, CompLex nr. 9/84. Universitetsforlaget, Oslo.
- Fjeldvig, Tove (1987) *Effektivisering av tekstsøke-systemer - utvikling av språkbaserte metoder*. Mastergradsoppgave ved Institutt for informasjonsvitenskap, CompLex nr. 7/87. Universitetsforlaget, Oslo.
- Fjeldvig, Tove/Golden, Anne (1987) *Automatisk splitting av sammensatte ord*. Vil bli publisert i rapportserien til NAVFs edb-senter for humanistisk forskning.
- Hansson, Hasse (1979) «Automatisk orddeling». Mægaard (1979) s. 33-42.
- Hellberg, Steffan (?) «Automatisk lemmatisering — En modell for opprättande av böjningsserier i ett frekvenslexikon». Göteborgs Universitet, Forskningsgruppen for modern svenska. Sprakdata. Göteborg.
- Jones, K.P./Bell C.L.M. (1984) «The automatic extraction of words from texts especially for input into information retrieval systems based on inverted files». van Rijsbergen (1984) s. 409-420.
- Karlsson, Fred (ed.) 1985 *Computational Morphosyntax — Report on Research 1981-85*. University of Helsinki. Publications No. 13 1985. Helsinki.
- Karlsson, Fred (1986) *Papers from the Fifth Scandinavian Conference of Computational Linguistics*. University of Helsinki. Publications No. 15 1986. Helsinki.
- Koskenniemi, Kimmo (1983) *Two-level Morphology: A General Computational Model for Word-Form Recognition and Production*. University of Helsinki. Publications No. 11 1983. Helsinki.

- Koskenniemi, Kimmo (1985a) «A General Two-level Computational Model for Word-Form Recognition and Production». Karlsson (1985) s. 1-18.
- Koskenniemi, Kimmo (1985b) «A System for Generation of Finnish Inflected Word-Forms». Karlsson (1985) s. 63-80.
- Källgren, Gunnel (?) «En algoritim för deling av sammensatte ord i svenska». Institutionen för lingvistik, Stockholm universitet.
- Maegaard, Bente (ed.) (1979) *Nordiske datalingvistikdage i København 9.-10. oktober 1979*. Institut for anvendt og matematisk lingvistik, Københavns Universitet. København.
- Niedermaier, G. m.fl. (1984) «MARS: a retrieval tool on the basis of morphological analysis». van Rijsbergen (1984) s. 369-382.
- Rankin, Ivan (1986) «SMORF — an implementation of Hellberg's morphology system». Karlsson (1986) s. 161-172.
- van Rijsbergen, C.J. (ed.) (1984) *Research and Development in Information Retrieval*. Proceedings of the third joint BCS and ACM symposium King's College, Cambridge 2-6 July 1984. British Computer Society Workshop Series.

Tove Fjeldvig er seniorkonsulent i Statens Datasentral, hvor hun arbejder i gruppen for informationsøking. Denne gruppen har ansvaret for vedlikehold, videreutvikling og markedsføring av fritekstsøkesystemet SIFT.

HUMANISTISK DATALOGI I DANMARK

Mogens Baumann Larsen

UDDANNELSESPLANLÆGNING I DANMARK

Tilfældigheder spiller en deprimerende stor rolle i danske universiteters planlægning af uddannelserne. Påvirkningsfaktorer som økonomi, uddannelsespolitik, landsplanlægning, fakultetspolitik, faglige strømninger blandt lærerne og studenterholdninger forårsager, at en eftertid har svært ved at finde frem til de påvirkninger, der lå bag det faktiske udbud af universitetsuddannelser på et givet tidspunkt. Det bliver eksempelvis en tung opgave for uddannelseshistorikerne at analysere forløbet af eksperimentet Roskilde universitetscenter.

På den baggrund fremstår det som en taknemmelig og enkel opgave at analysere faget Humanistisk datalogi. Faget opstod nemlig som en konsekvens af to klart formulerede behov. Det primære var et politisk krav om en indsats, der kunne skabe nye jobmuligheder for arbejdsløse humanister, og det andet var et rent fagligt ønske om, at den matematisk funderede datalogi fik et korrigerende supplement.

ARBEJDSLØSHEDEN

Allerede sidst i 70'erne stod det klart for alle planlæggere, at der ville komme en alvorlig arbejdsløshed blandt humanister i Danmark. Studenterne strømmede til de humanistiske fakulteter, og da ansættelsen af universitetslærere styredes automatisk af studentertallet, udvidedes de humanistiske lærerkollegier tilsvarende omend i en langsommere takt. Alle de nyansatte lærere fik forskningsret, så det stigende studentertal medførte også en massiv udvidelse af humanistisk forskning.

Selv om søgningen til humaniora steg hvert år, så blev de ungdomsårgange, der nåede frem til gymnasiealderen, stadig mindre, og dermed fjernedes det naturlige erhvervsgrundlag for de færdige humanister. Behovet for gymnasielærere reduceredes gennem 80'erne. Det kom så vidt i 1985, at gymnasielærere afskedigedes.

Alle danske planlæggere søgte trøst og inspiration i den angelsaksiske akademiske tradition, hvor der er en langt friere relation mellem

uddannelse og erhvervsfunktion. Magisteren i kinesisk, som bliver ansat i boligministeriet, var den forjættende vision. Det viste sig imidlertid hurtigt, at det er svært at indføre en sådan frihedsgrad i et statsapparat, der gennem århundreder har opbygget traditioner og aftaler om faglige kompetencer. Folkeskolen forblev også et lukket land for de arbejdsløse humanister.

HUMANISTER I ERHVERVSLIVET

Overraskende nok var det i erhvervslivet, de nye chancer for humanisterne opstod. Modvirkende kræfter fandtes også her i form af en ængstelse for den gøgeungevirkning, som ansættelsen af humanister ville skabe. Personale med ringere teoretisk baggrund ville blive forvist til mere rutinepræget arbejde og måske blive afskediget.

Dette og andre forbehold hindrede dog ikke, at det danske erhvervsliv i 80'erne ansatte et stigende antal humanister, og det sidste par år har det drejet sig om så mange, at det næsten svarede til antallet af kandidater, som blev færdigt i disse år.

Dette positive træk skyldes også, at virkningen af en stram adgangsbegrænsning til humaniora, som blev indført i 80'erne, nu slår igennem. Antallet af færdige kandidater er faldet stærkt de sidste år.

Arbejdsløshedsproblemet gik derfor over til at blive et restproblem. De 2-3000 kandidater, der ikke blev beskæftiget først i 80'erne, får stadig sværere ved at få arbejde, fordi erhvervslivet foretrækker de unge og nyuddannede kandidater. De unge har nu ofte fået en humanistisk uddannelse, der passer bedre til arbejdsopgaverne i erhvervslivet.

I efteråret 1987 har både undervisningsministeren og Arbejdsgiverforeningen udtrykt tvivl om, at den nuværende gode tilgang af humanister til erhvervslivet vil fortsætte. Den økonomiske stagnation vil i firmaerne først ramme de sidst og marginalt ansatte, nemlig humanisterne.

KAMPEN MOD HUM-ARBEJDSLØSHEDEN

I nært samarbejde med erhvervslivet (IBM og Handelsbanken bl.a.) gennemførte undervisningsministeriet en række suppleringskurser i datalogi for arbejdsløse humanister. Det drejede sig om intense fulltidskurser på 5 måneders længde, hvor lærere og pædagogisk tradition kom fra erhvervslivets personaletræning. Kurserne blev en succes, næsten alle gennemførte og fik stillinger i erhvervslivet.

Næste trin blev et ønske om, at universiteterne indenfor den normale humanistiske studieramme kunne give en tilsvarende kvalifikation.



Omskolert humanist, full av energi og optimisme. (Tegning: Øystein Reigem).

Resultatet blev, at der på Århus universitet i 1982 blev udbudt et humanistisk-datalogisk bifag på to år, hvor betingelsen var, at de studerende havde et gennemført humanistisk hovedfagsstudium bag sig eller en tilsvarende humanistisk uddannelse. I 1983 begyndte et hum-dat bifag på Aalborg universitetscenter.

Københavns universitet gik andre veje. Naturligvis ønskede universitetets planlæggere, at der skulle være mulighed for en datalogisk kvalifikation af humanistiske studerende, men hovedsynspunktet var, at det skulle ske gennem undervisningstilbud i det enkelte fag og uden det klare erhvervsigte. Forudsætningen var, at der efter en årrække skulle findes lærere i de enkelte fag, som havde fået en dobbeltkvalifikation gennem en supplerende datalogitræning.

Den langsigtede alvor i universitetets planlægning fremgår af, at der i efteråret 1987 er 15 humanistiske lærere, som er frigjort fra undervisningen for at gennemgå et intenst fuldtidskursus. Opkvalificeringen er arrangeret af en gruppe humanistiske kolleger, der gennem de sidste fire-fem år selv har fået mulighed for at opbygge den dobbelte kompetence.

Den københavnske løsning har indtil nu haft ringe gennemslagskraft på studenterniveau, og det vil også på lang sigt blive vanskeligt at vurdere, om de datalogiske kvalifikationer hos københavner-kandidaterne vil få betydning for jobmulighederne. Styrken ved løsningen er langsommeligheden. Den gradvise fagintegration mellem de enkelte humanistiske fag og datalogien vil givetvis skabe en dybtgående faglig fornyelse uanset reaktionen på arbejdsmarkedet.

HUM-DAT: DET FAGLIGE KONCEPT

En resonansvirkning opstod under planlægningen av hum-dat i Århus/Ålborg. *Peter Bøgh Andersen*, Nordisk institut, Århus universitet, havde som talsmand for en tværfaglig lærergruppe foreslået, at man skulle opbygge et datalogistudium med lingvistikken som videnskabeligt grundlag i stedet for matematikken. Dette forslag fik nu pludselig en ny status, fordi det motiverede og kvalificerede den ønskede job-kvalificering af humanisterne.

Indenfor lingvistikken var det ikke datalingvistikken, men disciplinen pragmatik, forslaget byggede på. Pragmatikken er indenfor dansk lingvistik en konsekvens af opgøret i 1968. Lærere og studerende på Københavns universitet skabte i 70'erne en dansk forsknings- og undervisningstradition i denne disciplin, hvor tekstanalysen fastholder afsender, modtager, medium, situationskontekst og samfund som aspekter. Da den nye datalogiske analyse ønskede at fastholde bruger og brugssituation som hovedaspekter, så var pragmatikken et naturligt sproganalytisk grundlag.

På samme måde som pragmatikken opstod i et samspill mellem sprogvidenskab, filosofi og socialpsykologi, blev hum-dat et samspill mellem disse fag og datalogien.

Uanset den massive indflydelse fra lingvistikken var hum-dat baseret på det datalogiske. Hum-dat skulle primært arbejde med computerbaserede systemer og menneskers rolle ved udvikling og anvendelse af disse systemer. Den norske datalog *Kristen Nygård* er således en forudsætning for hum-dat gennem interessen for brugerne (samarbejdet med fagbevægelsen) og det åbne faglige koncept. Et Nygård-citat fra 1986 viser, at Nygård har en fagopfattelse, der ved sin totalitet giver rigelig plads til en fornyelse som hum-dat: «Informatics is the science that has as its domain information processes and related phenomena in artifacts, society and nature» (fra *Program Development as a Social Activity*, 1986).

Systemanalysen blev det centrale i hum-dat, og den erhvervs-kvalifikation, som studiet skulle føre frem til, var en utraditionel og kritisk tilgang til systemarbejdet.

KVINDER OG EDB

Det nye rekrutteringsgrundlag for studerende måtte også føre til en ændring i forhold til de klassiske datalogimiljøer, idet der er flest kvinder blandt de humanistiske hovedfagskandidater. Temaet: kvinder og edb blev også centralt for hum-dat og fik umiddelbare konsekvenser uden for studiemiljøet, fordi kvindelige studerende med stor tilslutning førte debatten videre på Folkeuniversitetet om dette emne.

HUM-DAT: PLACERING PÅ UNIVERSITETET

Strukturelt opstod det samme problem for hum-dat, som samtidig var opstået ved planlægningen af et andet nyt felt: områdestudierne (area studies). Studier på tværs af fag og fakulteter kan organiseres som et centerbaseret samarbejde mellem forskellige institutter og fakulteter. Den anden mulighed er, at det ny studium organiseres ved oprettelse af et selvstændigt institut, der inden for sin ramme samler de nødvendige faglige kompetencer.

I Århus valgte man samarbejdsmodellen, idet hum-dat blev en fælles aktivitet for det naturvidenskabelige og det humanistiske fakultet. I Ålborg ansatte det humanistiske fakultet en datalog, *Inger Lytje*, til at opbygge hum-dat inden for fakultetets ramme. Ålborg-strukturen har givetvis været den, der fungerede bedst, men de administrative forskelle har været af underordnet betydning for hum-dat-udviklingen.

HUM-DAT: DE FØRSTE ÅRGANGE

I planlægningsfasen var der kontakt mellem universitet og erhvervsliv. Planlæggerne i Århus bad en række firmaer om en vurdering af erhvervs-mulighederne for de kommende hum-dat kandidater. Undersøgelses konklusion var, at der ville blive gode erhvervs-muligheder for hum-dat. De følgende år viste, at prognosen var rigtig, for alle de uddannede kandidater fik job.

Set fra en eftertid er det faktisk overraskende, at man kom frem til en sikker prognose, for uklarhed prægede universitetets oplysninger om, hvad bifagskandidaterne ville kunne, og vaghed prægede erhvervs-livets redegørelser for de kvalifikationer, man savnede. Selv om der

var en beskeden praktik indlagt i studiet, var der ikke megen kontakt mellem studium og edb-branche, ligesom man heller ikke knyttede deltidslærere fra erhvervslivet til studiet. Hum-dat forblev en universitetsaktivitet.

De første årgange af studerende delte planlæggernes ønske om at lave en uddannelse, der på samme tid var en faglig fornyelse og en forbedring af jobchancen. De første årgange var tilfredse med den valgte arbejdsform: gruppearbejde og projektorientering, og den til tider voldsomme diskussion mellem lærere og studerende om studiets indhold blev af begge parter opfattet som en velindarbejdet tradition fra 1968. Lærere og studerende var dog enige om at fremhæve det kritiske potentiale i studieplanerne, og teknologikritikken blev en vigtig disciplin. Det viste sig at være svært at finde frem til en indlæringsform for studiets færdighedsmomenter, og de krav, som beherskelse af selv elementær programmering stiller, var et tidkrævende chok for humanisterne.

I 1984 blev de første kandidater færdige og fik stillinger. Hverdagen var ved at blive konsolideret på hum-dat. Men hverdagen rummede en stiltfærdig forandring. Forbedringen av erhvervschancen blev et mere dominerende perspektiv for de nye årgange, og fagets kritiske moment interesserede mindre. Samtidig skete der en stiltfærdig forskydning i det pædagogiske, fordi flere og flere studerende valgte at arbejde individuelt.

NY HUMANIST-STRUKTUR PÅ UNIVERSITETERNE

I princippet skulle den nye humaniststruktur, der kom i 1985, været helt uproblematisk for hum-dat. En grundide ved reformen var, at det skulle være muligt for den studerende at kombinere fagmomenter, også momenter uden for humaniora, så den afsluttende kvalifikation kunne tilpasses erhvervslivets behov. Planlæggerne havde netop tænkt på fagmomenter fra erhvervsøkonomi og datalogi som ønskværdige for humanisterne.

Både i Århus og Ålborg valgte man at lade det to-årige bifagsstudium integrere i nye fagdannelser. I Århus gled hum-dat organisatorisk og fagligt sammen med disciplinen massekommunikation, og man fik nydannelsen: Informationsvidenskab. I Ålborg blev hum-dat en specialisering i sidste del av studiet ved siden af kommunikation, og den samlende ramme er nydannelsen: Humanistisk informatik. I stedet for en to-årig overbygning på et humanistisk hovedfag fik man to fireårige specialiststudier.

HUM-DAT: ET FORSVUNDET PERSPEKTIV

I undervisningsministeriet havde man med tilfredshed fulgt udviklingen af hum-dat. Forslagene om de nye fag: Informationsvidenskab og Humanistisk informatik fik ligeledes en positiv modtagelse, for også denne gang drejede det sig om nye fag, der var blevet planlagt ved et samarbejde mellem universiteter og erhvervsliv. Der var en forsigtig enighed om, at den nye type kandidater nok skulle finde arbejde.

Det ejendommelige var, at hum-dat, som var en fastslået succes, var på vej ud efter fem år. Humanister med supplementet hum-dat var placeret og fungerede til alles tilfredshed, og man havde velfungerende hum-dat-studier, men alligevel nedlagde man disse. En kyniker vil sige, at en bedre skæbne kan et studium ikke få. Den bratte og umotiverede afslutning kan dog få et harmonisk perspektiv, hvis de nye faglige sammenhænge viser sig at være endnu mere kreative.

Tilbage bliver det faktum, at hum-dat var en god ide, der forsvandt. Mens hum-dat kandidaterne var generalister, bliver de nye kandidater eksperter, som kommer til at konkurrere om stillingerne med andre eksperter: journalister og pr-folk.

Hum-dat blev en kortvarig succes, og vi er en lille gruppe, der allerede savner den.

Mogens Baumann Larsen er professor ved Aalborg universitetscenter og formand for det faglige landsudvalg for humaniora, undervisningsministeriet.

RAPPORTER

INTERAKTIV VIDEO I FREMMEDSPRÅKS- UNDERVISNINGEN

Reidun Oanæs Andersen

Høsten 1986 ba Datasekretariatet meg gjøre en undersøkelse om bruk av interaktiv video (IV) i fremmedspråksundervisningen, fortrinnsvis fransk. Mandatet gikk ut på følgende: innhente informasjon om ferdige eller prosjekterte IV-program til bruk i fremmedspråksundervisning, demonstrere og eventuelt prøve noen av disse programmene ut i fremmedspråksmiljøene omkring Pedagogisk Seminar i Bergen og rapportere tilbake til Datasekretariatet med evaluering av programvaren samt funksjonsanalyse. I det følgende skal jeg presentere en del IV-program for fremmedspråk og vurdere disse ut fra Læreplanens syn på fremmedspråksundervisning.

Under en workshop på Centre International d'Etudes Pédagogiques i Sévres 1986 fikk jeg for første gang demonstrert et IV-program beregnet for fremmedspråksundervisning. Det gjorde inntrykk. Jeg så at en ved hjelp av moderne teknologi kunne få mulighet til å arbeide med helheten i en kommunikasjonssituasjon. Det betyr en radikal forbedring i forhold til de fleste eksisterende dataprogram som er atomistiske i sin struktur. De fokuserer på grammatiske korrekte setninger, ofte løsrevet fra sin sammenheng, slik at budskapet i det språklige utsagn blir av mindre betydning.

Dagens kommunikative fremmedspråksundervisning er opptatt av at det språklige utsagn har en intensjon og skal formidle et budskap, det være seg i en litterær tekst eller gjennom en dialog. Likeledes at

utsagnet står i en sosial og kulturell sammenheng. Læreplanen for den videregående skole har kommunikativ kompetanse som overordnet målsetting og «Un Nouveau Seuil» definerer språk og språkbruk som «une action sociale». Ytringens form preges av kommunikasjonssituasjonen, hvilket budskap som skal formidles til hvem og gjennom hvilket medium. Rammene omkring det som sies eller skrives, blir derfor viktige. Det sosio-kulturelle aspekt likeså.

Å ha kommunikativ kompetanse er noe langt mer enn å kunne produsere setninger som er grammatisk korrekte. Språk og språkbruk er en del av et lands kultur. Evnen til å forstå og vise innsikt i sosio-kulturelle forhold knyttet til målpråket, blir på den måten viktige komponenter i tilegnelsen av et fremmedspråk.

Ut fra dette synet på fremmedspråk nærmet jeg meg de nye IV-program. Ville disse være en hjelp i arbeidet med å realisere noen av de målene som er nevnt ovenfor?

TEKNOLOGIENS MULIGHETER

La oss repetere en del karakteristika ved dette mediet:

- 54.000 bilder kan lagres på hver side av en videoplate som er 30 cm i diameter. Platen leses ved hjelp av en tynn laserstråle, og platen slites ikke ved bruk.
- Hvert bilde kan «fryses,» og en kan spole hurtig eller langsomt frem og tilbake.
- Hvert bilde har et nummer. En kan «hoppe» til hvilket som helst av de 54.000 bildene i løpet av noen sekunder. Studenten er derfor ikke bundet av en kronologisk avlesing av platens innhold.
- Informasjonene som ligger på platen, kan være av ulik art: stillbilder, levende film, stemmer, musikk, tekst.
- Videoplaten har to lydkanaler.

Hele denne banken av bilder og andre informasjoner kobles til en datamaskin og kan på den måten styres av et dataprogram. En øyner at denne kombinasjonen av tekst, lyd og bilde har store pedagogiske muligheter, ikke minst i fremmedspråksundervisningen. Et slikt medium kan langt på vei gjenskape en autentisk kommunikasjonssituasjon, noe som ofte er vanskelig i vanlig undervisning i fremmedspråk. Bilder og illustrasjoner kan gi rammene omkring det språklige utsagn og samtidig gjøre forståelsen lettere, både når det gjelder innhold og kontekst.

Mimikk og kroppsspråk regnes i dag som viktige elementer i en

autentisk muntlig kommunikasjonssituasjon. I arbeidet med disse sidene av fremmedspråkforståelsen kan kombinasjonen av lyd og bilde gi større innsikt og forståelse.

Gjennom styringsprogrammet kan det legges inn hjelpefunksjoner som ett- eller tospråklig ordbok, undertitler, forklaringer av grammatiske og leksikalske vanskeligheter. Disse mulighetene kombinert med mediets fleksibilitet når det gjelder repetisjon setter studenten i stand til å arbeide selvstendig og gjøre valg, i pakt med egne behov. Dermed får han ansvar for viktige sider ved egen læring.

UTVIKLING AV PROGRAMVARE

På denne bakgrunn er det interessant å se hvordan mediet foreløpig er utnyttet i noen av de IV-program som foreligger. La oss med en gang si at disse er ikke mange. Storbritannia og Frankrike driver forsøk med IV i skolen. Teknisk utstyr er plassert i et stort antall undervisningsinstitusjoner, programvare er under utvikling og utprøving, og det arrangeres kurs for lærere. Andre land er foreløpig ikke kommet så langt, og bare noen få har produsert sine egne IV-program. Dette har flere forklaringer. Et IV-program er dyrt å produsere, både fordi filmopptak og pressing av selve platen er kostbar, men også fordi et prosjekt av denne art må baseres på teamarbeid mellom eksperter på ulike felt. Den pedagogiske design og oppbygging av programmet er et krevende arbeid, både når det gjelder tid og ressurser. Men når platen først er presset og styringsdisketten er ferdig programmert, er det forholdsvis billig å lage mange kopier.

Av de prosjektene som er satt i gang i Storbritannia og Frankrike når det gjelder IV i fremmedspråksundervisningen, er noen avsluttet, andre ikke. Men felles for alle prosjektene er at programvaren blir betraktet som et eksperiment og således ikke er tilgjengelig på det åpne marked. Det er også vanskelig å få tak i programvaren for utprøving og demonstrasjoner. Det finnes IV-program i handelen, men få av disse er spesielt beregnet på fremmedspråksundervisning.

Men denne situasjonen vil etter alt å dømme forandre seg i nær fremtid da det synes å være en eksplosjonsartet utvikling på dette feltet akkurat nå.

PROSJEKTER

«Peau d'Ane»

På Expolangues 1985 presenterte Le Centre Mondial Informatique ved

direktøren *Georges Broussaud*, det første IV-prosjektet for fransk som fremmedspråk i Frankrike: «*Peau d'Ane, vidéodisque interactif de français, langue étrangère*». Det er beregnet på engelsktalende studenter, nivå 2 eller 3 og hensikten med prosjektet var å prøve ut IV-teknologiens muligheter. For å spare utgifter til opptak av eget bildemateriale, benyttet prosjektet seg av allerede eksisterende levende film og stillbilder, da først og fremst den franske spillefilmen «*Peau d'Ane*». I tillegg ble det lagt inn en del reklamefilmer, sportsklipp, telefonsamtaler og stillbilder som spesielt egnet seg som utgangspunkt for fremmedspråksinnlæring. Denne fremgangsmåten sparte ressurser, men innskrenket samtidig muligheten til fri vandring i en stor billedbank som er en av de store fordeler ved dette mediet. Skal muligheten til fri vandring være tilgjengelig, må billedmaterialet spesielt tilrettelegges for dette formålet.

En av ideene med prosjektet var å gjøre programmet og utstyret så brukervennlig som mulig. Studenten styrer programmet ved hjelp av direkte berøring av symboler på skjermen eller ved hjelp av mus. Tekstene fra styringsprogrammet legges på bildeskjermen.

Når det gjaldt den pedagogiske oppbygging av programmet, lot prosjektgruppen seg inspirere av Piagets prinsipper for språkinnlæring. Disse kan kort oppsummeres slik: Barnet lærer språk gjennom sterk påvirkning, utvelgelse, imitasjon og rekonstitusjon. Gruppen mente at disse prinsippene også måtte gjelde den voksne språkstudent og ville gjennom programmet gi studenten anledning til å oppleve et «lingvistisk og kulturelt bad» der språkbruken lå opp til naturlig kommunikasjon.

Under arbeidet med å forstå og selv produsere på fremmedspråket kan studenten benytte seg av ulike hjelpefunksjoner: velge ulike sekvenser, repetere x ganger visse passasjer, hente inn undertitler på fransk eller engelsk, lese en lettere språklig versjon av replikkene, se hele teksten til sekvensen og bla i denne som i en bok, hente inn grammatiske og leksikalske forklaringer og også se dem illustrert i andre kontekster.

Jeg har sett «*Peau d'Ane*» demonstrert, men ikke fått anledning til selv å prøve det ut på grunn av problemer i forbindelse med filmrettighetene. Programmet er omfattende og fleksibiliteten er stor, rent metodisk. Slik er det et godt eksempel på det potensiale som ligger i selve teknologien. Men innholdsmessig og pedagogisk har programmet store begrensninger, fordi en har tatt utgangspunkt i allerede eksisterende materiale som er spilt inn for andre formål. En er da bundet av de situasjoner og det språk som ligger ferdig innspilt, og må innrette

det videre arbeid etter dette.

Selv om en kan se verdien av å bygge et IV-program rundt et godt filmmateriale, må en kunne stille spørsmål ved riktigheten av å bruke hele eller deler av kunstverk utelukkende til arbeid med praktisk språkbruk. En spillefilm av denne type er en helhet. Skal den danne utgangspunkt for et pedagogisk arbeid, burde ulike sider ved kunstverket taes opp, også de kunstneriske. Slik kunne et slikt verdifulle materiale danne utgangspunkt for en dypere språkforståelse. Vil en bare konsentrere arbeidet omkring ferdighetstrening, vil andre typer filmer som dokumentarfilmer og reklamefilmer egne seg bedre.

IVIS-prosjektet

I april 1987 ble THE INTERACTIVE VIDEO IN SCHOOLS PROJECT (IVIS) lansert i Storbritannia. The Department of Trade and Industry har finansiert utviklingen av 8 IV-program som prøves ut i 90 ulike skoler dette året. Hele prosjektet koster 1.5 millioner pund.

Hensikten med prosjektet er å undersøke IV's muligheter som pedagogisk hjelpemiddel og å utforme programvare med stor fleksibilitet som kan møte både læreres og elevers behov. *Bill Plummer*, en tidligere rektor, er leder for prosjektet som administreres fra NIVC (The National Interactive Video Centre, London). Medlemmene i prosjektgruppen er erfarne lærere. De blir supplert av audio-visuelle eksperter og programmere, når dette er nødvendig. Programvaren skal kontinuerlig utvikles og forbedres og lærernes synspunkter og erfaringer ved bruk av programvaren, blir svært viktig i dette arbeidet.

Prosjektet omfatter følgende fagområder: miljøstudier (o-fag), geografi, design, naturfag, lærerutdanning, praktisk matematikk, fransk og personlighetsutvikling. Sentrene for de enkelte IV-program er spredt over hele Storbritannia, det samme er forsøksskolene. Således ligger senteret for praktisk matematikk ved Exeter University, mens franskprogrammet er lagt til The Shropshire Education Authority.

Programvaren tilhører prosjektet og ansees ikke som ferdig utviklet så lenge forsøket pågår, derfor kan en bare få kjennskap til programvaren gjennom demonstrasjoner og foredrag og gjennom et begrenset skriftlig materiale.

THE SHROPSHIRE INTERACTIVE VIDEO PROJECT er et IV-program for fransk som fremmedspråk, beregnet på elever fra 9-11 år. Målsettingen er å utprøve mediets potensiale i fremmedspråkundervisningen gjennom simulering av én serie autentiske situasjoner. Med vilje har de unngått drilløvelser.

Eleven føres inn i en fransk by og deretter i en gate der han kan vandre rundt ved hjelp av kurseren. Han disponerer en viss pengesum og skal bruke denne til å gjøre innkjøp i ulike butikker. Han må svare på spørsmål fra en politimann og forskjellige ekspeditører. Elevens svar etterfølges alltid av en konsekvens som står i forhold til det svaralternativ eleven har valgt.

Filmene og illustrasjonene til dette IV-programmet er spilt inn spesielt for formålet og er av høy teknisk kvalitet. Rammene omkring selve kommunikasjonen er optimale. En fransk grønnsakhandler rekker smilende fram et eple og spør hvor mye vi vil ha. Dette skulle vel være en tilnærmet autentisk situasjon så god som noen. Men hvorfor er den ikke det? Fordi grønnsakhandleren og bakeren — ja, alle personene i den franske gaten — ikke fører et språk som er naturlig i situasjonen. De snakker som en lærebok. Hva verre er: De snakker i multiple choice varianter. Eleven blir heller ikke utfordret til å produsere språk, bare trykke på et nummer. Så snakker en stemme på platen i stedet. Elevene kan gjøre spennende vandringer inn og ut av ulike butikker, men den pedagogiske ide for fremmedspråksinnlæring som ligger til grunn for programmet, må en stille seg noe spørrende til. En har inntrykk av at teknologien har fått bestemme hvilke pedagogiske løsninger som skulle velges, ikke det motsatte, som skulle være det eneste riktige i denne sammenheng.

The North West Educational Computing Project

Dette prosjektet har en helt annen innfallsvinkel til sine IV-program i fransk og tysk. Med støtte fra britisk IBM og i samarbeid med lokale skolemyndigheter og fire skoler i Lancashire har programmereren *Ian Robertson* meget langt på vei klart å realisere den erfarne språklærer *Mike Picciottos* intensjoner og pedagogiske ideer. Picciotto er ikke interessert i at studenten skal avgi det ene rette svar som skal plottes inn i maskinen. Hans hovedintensjon er å få studenten til å produsere naturlig muntlig fransk og tysk gjennom å presentere mennesker i autentiske situasjoner som snakker om interessante ting. Han vil utnytte mediets muligheter slik at studenten får den hjelp han trenger. Studenten blir ikke styrt gjennom en på forhånd uttenkt pedagogisk progresjon. Programmet er åpent. Studenten inviteres til å gjøre egne valg av arbeidsoppgaver ut fra nivå og behov og til å samspille med maskinen.

Gruppen har gjort en bearbeiding av BBC-platen ENSEMBLE. Men deres egne program omfatter en rekke dialoger som er spilt inn i

skolemiljø i Frankrike og Tyskland der personene snakker et naturlig språk i naturlig hastighet om emner som unge mennesker kan identifisere seg med. Her er ingen stereotypisk visitt hverken på postkontor eller i bank, men samtaler som engasjerer på det personlige og følelsesmessige plan. Dette autentiske materialet danner rammen omkring et vell av ulike arbeidsoppgaver og hjelpefunksjoner, der målet er at studenten skal bli så fortrolig med kommunikasjonssituasjonen og det muntlige språk at utsagnene nesten blir en del av ham selv. Han kan gå inn som en av partene i dialogen. Er en kommet så langt, er ikke veien lang til fri språklig produksjon.

Disse programmene fokuserer også på de ekstra lingvistiske elementer som mimikk og kroppsspråk og på deres betydning i kommunikasjonssituasjonen både som formidlere av senderens budskap og som hjelp til forståelse for språkstudenten.

Begrensningen i disse programmene ligger i valg av kommunikasjonssituasjoner og emner. Disse må aksepteres i utgangspunktet, og en diskusjon omkring hvorvidt emnene er sentrale og viktige i forståelsen av fransk språk og miljø, er selvfølgelig berettiget.

En må likvel si at dette prosjektets utradisjonelle tilnæringsmåte til mediet kombinert med respekten for brukerens frihet til å ta egne pedagogiske avgjørelser gjør at disse programmene stiller i første rekke når det gjelder en utnyttelse av mediet som samsvarer med den overordnede målsetting for fremmedspråksundervisningen. Interessen og bruksfrekvensen blant elevene i forsøksskolene tyder på at dette må være rett. Heller ikke disse programmene er for salg, da de er prosjektets eiendom. Men platene ble demonstrert på et kurs på Vestlandske Lærarstemna i Bergen 9. okt. 1987 og på konferansen «TEKST, LYD, BILDE» som NAVFs edb-senter for humanistisk forskning organiserte i Bergen 14.-16. okt. 1987.

Gruppen arbeider videre med utnyttelse av IV-teknologi i arbeidet med språkfunksjoner og register. Ideen er å bruke bildene aktivt for å lette forståelsen av den kommunikative intensjon og vise hvordan budskapet kan få ulik form etter hvilket sosialt miljø de talende befinner seg i. (Se også Robertson og Picciottos artikkel i dette nr. av HD. Red.)

Avec Plaisir

Det franske forlaget HACHETTÉ har i flere år arbeidet på et IV-program i fransk som fremmedspråk: «Avec Plaisir». Utgivelsen har

vært sterkt forsinket, og bare deler av programmet er vist på workshops og konferanser. Innholdet på platen er imidlertid kjent. Deler av video-tapen «Avec Plaisir» er lagt over i bearbeidet form, slik at IV-teknologien utnyttes. Platen inneholder også flere hundre reklameplakater, klipp fra reklamefilmer og nyhetssendinger på fransk fjernsyn, samt en film om komponisten Berlioz. Flere av reklamefilmene er utnyttet for å gi studenten anledning til å forstå og å arbeide med ulike språkfunksjoner.

Noen vurdering av «Avec Plaisir» er ikke mulig, da jeg bare har sett små utsnitt av programmet. Det skal komme i salg i høst til en pris av ca. 2000 FF. En kan imidlertid understreke det interessante i at Hachette lager et IV-program som presenterer ulike sider ved landets kultur, ikke bare språket. Innblikk i kunst og kultur, kjennskap til aktuelle samfunnsproblemer og «documents authentiques» øker innsikten og motivasjonen hos studenten og setter språkstudiet inn i en større og mer meningsfull sammenheng.

Voyage en français interactif

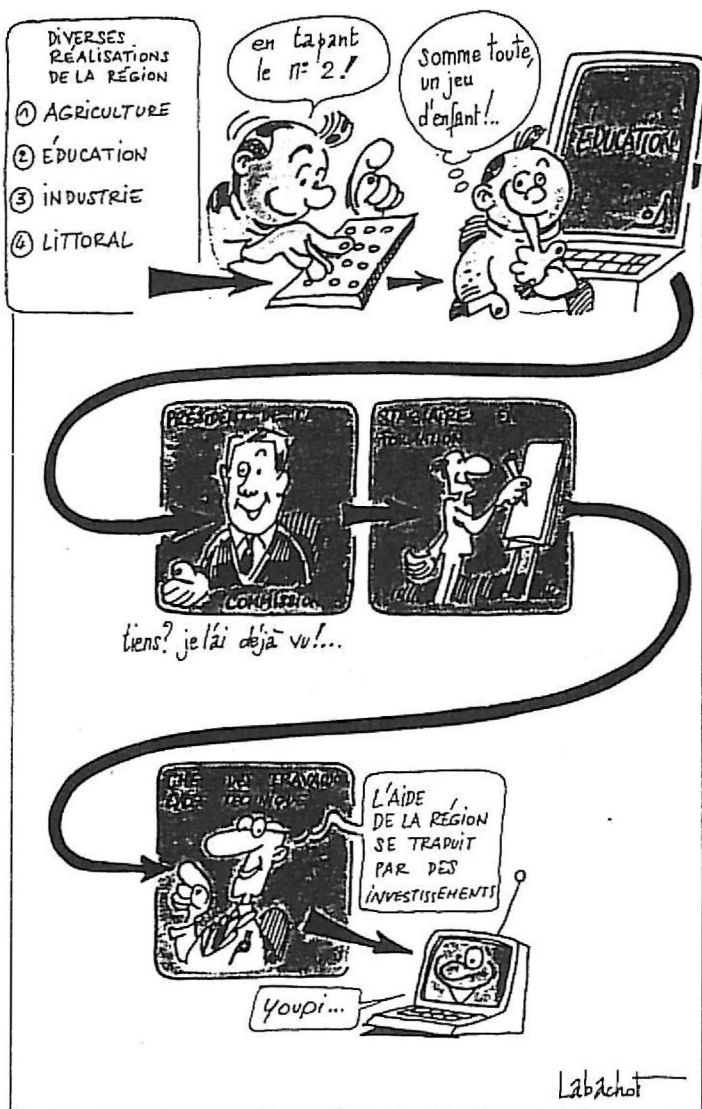
Forlaget NATHAN arbeider også på et IV-program i fransk som fremmedspråk beregnet på ungdom og voksne som har litt kjennskap til fransk. Programmet skal inneholde en blanding av tegneserier, leker, autentiske situasjoner og tester. Men da jeg ikke har hatt anledning til å se programmet, må en nærmere vurdering av innhold og pedagogisk ide vente til platen foreligger på det åpne marked.

Andre IV-program

Videoplate-gruppen ved ACADEMIE DE POITIERS, det regionale pedagogiske senter for videregående opplæring har til nå utviklet 3 IV-program: ett om billedanalyse, ett om regionen Poitou-Charentes og ett om yrkesorientering. Programmene er resultat av utprøving av nye medier og et ledd i arbeidet med å gi elevene større frihet og dermed ansvar for egen læring.

Programmet om yrkesorientering bruker bildene som innfallsvinkel til de ulike avsnittene og emnene, mens programmet om regionen er en stor tekst-, lyd- og billedbase som gir opplysninger om regionens administrasjon, folkeliv og særpreg.

Disse programmene er beregnet på elever i den videregående skolen i Frankrike. De viser ulike mennesker i autentiske situasjoner som snakker om sitt yrke, sitt dagligliv eller sine politiske holdninger. Det



Her reklameres det for det nye IV-programmet som Académie de Poitiers har laget om geografi og samfunnsliv i regionen Poitou-Charentes. 200 gymnas har IV-utstyr og kan arbeide med dette og andre program.

synes opplagt at en slik rikdom av språklige, sosiale og kulturelle impulser burde utnyttes i undervisningen av fransk som fremmedspråk. Det språklige nivå er en barriere for språkstudenten. På den annen side vil rammene omkring de ulike samtalene, kroppsspråk og mimikk i stor grad lette forståelsen.

Det samme gjelder THE ECODISC som er et biprodukt av det store Domesday-prosjektet, laget av BBC. Det er et simuleringssprogram der studenten utfordres til å ta stilling til miljøspørsmål i et naturreservat. Alle avgjørelser studenten tar, får konsekvenser, positive eller negative. Hensikten med programmet er å gi den engelske student innsikt i den kompliserte og viktige problematikk som er knyttet til et slikt emne. Men det samme programmet kan gi den utenlandske språkstudenten innsikt i engelske samfunnsspørsmål og engelsk geografi. Og fordi disse spørsmålene er engasjerende og programmet er godt bygget opp, kan det øke hans forståelse av ord og uttrykk som hører hjemme i denne sammenheng og motivere ham for videre læring.

Fleire franske museer som Centre Georges Pompidou, La Villette og Musé d'Orsay har lagt store mengder av sitt innhold på videoplater, slik at brukerne kan foreta simulerte vandringer i museet og innhente flere opplysninger om spesielle objekt som interesserer dem. Annen informasjon i relasjon til museets spesielle oppgaver er også tilgjengelig på IV.

En ville ønske at fremtidens IV-program for fransk, som kanskje skal brukes langt borte fra Frankrike, ville inneholde ikke bare materiale for arbeid med språk, men også gi bakgrunnskunnskap gjennom presentasjon av samfunnsspørsmål, kunst og kultur. Da vil en kunne si at programmene virkelig ville være en hjelp for å arbeide i samsvar med Læreplanens intensjoner.

IV-teknologien åpner for uante muligheter når det gjelder kombinasjon av lyd, tekst og bilde og formidling av kunnskap, forståelse og innsikt. Alt dette kan utnyttes i fremmedspråkundervisningen. Men en må nærme seg det nye mediet ut fra et pedagogisk ståsted. Mediet bør brukes, ikke bare som en demonstrasjon av teknologiens muligheter, men fordi dette i noen tilfeller vil være den beste måten å arbeide på.

Reidun Oanæs Andersen er lektor med fransk som hovedfag. Hun er tilknyttet Pedagogisk seminar i Bergen som fagmetodikk lærer i fransk. For tiden har hun permisjon for å utvikle en test i norsk for fremmedspråklige ved Avdeling for fremmedspråklige studenter, Universitetet i Bergen. Hun er også engasjert på deltid ved NAVFs edb-senter for humanistisk forskning for å arbeide med interaktiv video.

RD-BIB

BIBLIOTEKSYSTEM FOR MIKROMASKINER

Eli Schilbred

I *Synopsis* vol. 118(1987): 1 presenterte Rogalandsdata A/S sitt biblioteksystem for mikromaskiner: RD-BIB.

Ved henvendelse til RD fikk vi opplyst at de ikke hadde noen demoversjon for testing, men NAVFs edb-senter for humanistisk forskning kunne få tilsendt systemet med innlagt datosperre, testperioden ble 3 mnd.

Formålet med testingen var dels at man ønsket å gjøre seg kjent med systemet, og dels å teste det for eventuelt bruk i biblioteket på Statens høyskole for kunsthåndverk og design, SHKD (Bergens kunsthåndverksskole).

SHKD kjøpte systemet i juni 1987, og det følgende er opplysninger og erfaringer basert både på testperioden og på bruk av systemet etter oppstart på SHKD.

TEKNISKE FORUTSETNINGER

Minimumskrav er en IBM-kompatibel PC med 20 MB harddisk og 1 skriver. Operativsystemet er MS-DOS. — Programmet alene krever 1.5 MB lagringsplass.

Det er ikke noe krav at man må ha en AT, men jeg vil tro at det vil være en merkbar svakhet å kjøre systemet med en langsommere maskin, fordi det etter hvert vil bli søking i større og større datamengder og svartiden kan bli lang.

I presentasjonen av RD-BIB som man kan få ved henvendelse til firmaet, er det ikke nevnt noe om spesielle krav til maskinens internminne. I brukerdokumentasjonen, som man altså ikke får før man har kjøpt systemet, opplyses det at internminnet må være minimum 256K, men at systemet er utviklet på en maskin med internminne 640K.

På SHKD, som har en maskin med internminne 512K, fikk vi imidlertid feilmelding i én funksjon: funksjon 45 **Tittelsletting**. Vi sendte en systemdiskett til RD og ved testing fikk også de feilmelding på

maskiner med internminne på 512K, mens systemet virket tilfredsstillende på maskin med 640K.

Maskinen på SHKD kan oppgraderes, men dette betyr jo en merutgift som vi ikke hadde regnet med. Dette bør komme med i RDs generelle presentasjon.

Systemet er tenkt brukt i små bibliotek med anslagsvis 1000 titler, og i mellomstore bibliotek med oppimot 40.000 titler.

For sikkerhetskopiering/backup kan man bruke disketter, men etter hvert som datamengden øker, vil dette bli tungvint. RD anbefaler streamer, men her må man være oppmerksom på at prisen på maskinvarer øker med opp til 10.000 kr.

Systemet kan brukes som et énbrukersystem, eller det kan kobles opp i et lokalt nettverk med flere arbeidsstasjoner og felles lagringsenhet.

Man kan også etablere tilknytning til eksterne system for søking, og det er mulig å få utskrift av søket, men man kan ikke hente inn katalogposter til RD-BIB. MARC-formatet er ikke brukt.

Systemet er rimelig. RD opererer med to priser. Skolebibliotek: Kr 7.500. Folke- og fagbibliotek: Kr 15.000. Prisene er ekskl. moms.

GENERELT OM SYSTEMET

Brukerveiledningen er en velordnet ringperm, med stort sett greie forklaringer, unntatt er funksjon 98 **Rapportgenerator**. Det er tydelig at man har lagt vekt på at det ikke skal være nødvendig med særlige edb-kunnskaper for å kunne bruke systemet.

Oppstart og innlegging er det fullt mulig å klare på egen hånd, hvis man tar seg litt tid og følger brukerveiledningen nøye.

Klassifikasjonssystemet er Dewey. Jeg har ikke kjennskap til om det ville være mulig å bruke f.eks. UDK, eller andre klassifikasjonssystemer.

Systemet gir anledning til å definere 5 brukergrupper i tillegg til brukergruppe A som er definert av systemet til å ha adgang til alle funksjoner. Det er altså mulig å gi lånere adgang til selv å bruke systemet til tittelsøking, uten at de har muligheter til å endre data. Assistenten i biblioteket kan få adgang til å bruke en del av funksjonene, slik som purring, utlån, reservering o.l. I praksis må man nok ha mere enn én terminal hvis lånere skal kunne søke selv.

Systemet er bygget opp i en serie faste funksjoner. Ill. 1 viser hovedmenyen. Man kan gå ut og inn i de forskjellige funksjonene og veksle mellom innlegging av dokumenter og service-funksjoner som utlån,

Rd-Bib - Biblioteksystem utviklet av Rogalandsdata A/S		
Funksjoner	Systemvedlikehold	Utskrifter
01 Tittel	30 Statistikk	50 Etiketter
02 Søkeord	31 Kodeverk	52 Forfalte lån
03 Eksemplar	36 Postkatalog	
06 Låner	45 Tittelsletting	
07 Utlån		
08 Innlevering		
09 Reservering		
10 Reserverte titler		
20 Søkning	96 Eksport	
	97 Import	
	99 Systemparametre	98 Rapportgenerator

F1: Funksjon + Sek ESC: Avslutt Rd-Bib

III. 1. Hovedmeny.

reservering etc. uten problemer. — Betingelsen er bare at man oppdaterer med en funksjonstast og så bekrefter oppdateringen med samme funksjonstasten, før man går inn i en ny funksjon, ellers mister man innlagte data i siste skjerm bilde man har arbeidet i.

Hvert skjerm bilde har nederst en oversikt som viser bruk av funksjonstastene, slik at det ikke hele tiden er nødvendig å lese i bruker-veiledningen når man gjør seg kjent med systemet.

POSITIVE SIDER

Systemet er stort sett enkelt å installere, enkelt å lære og enkelt i bruk.

Det har øyeblikkelig oppdatering, noe som virkelig forenkler arbeidet, men som samtidig er nokså nødvendig når det skal ta seg av både dokumentregistrering, utlån, reservering, lånerregister, bestandsinformasjon, statistikk og rapporter.

I funksjon 31 kan man bygge opp sitt eget kodeverk i den grad det er nødvendig for det enkelte bibliotek. Systemet gir mulighet for inndeling i genre, lesergruppe, media, nivå og språk.

På SHKD har vi foreløpig bare brukt kodeverket til å definere media og språk. Ill. 2 viser SHKDs oppdeling av media. Vi skiller hos oss mellom bøker, årbøker, utstillingskataloger og konferanserapporter. Men systemet gir også mulighet til å registrere video, film, kassetter med musikk eller plater. Der er egne felt for spilletid og et telleverk hvor man kan angi start og stopp på videogrammet, filmen etc.

RD-Bib #
Funksjon:31 Sak:
Melding :

31 Kodeverk:

06/11-1987 10:57

Hovedgruppe: M

Gyldige grupper er G for genre
L for Lesergruppe
M for Media
N for Nivå
S for Språk

OPPDELING AV HOVEDGRUPPE

Kode	Betegnelse
B	bok
K	konferanserapport
U	utstillingskatalog
__A	(ikke spesifisert)
A	årbok

F1:Funksjon+Søk F3:Slett F5:Les neste F7/F8 :Forrige/Neste side ESC:Home
F2:Blank bilde F4:Les F6:Oppdater F9/F10:Første/Siste side

III. 2. SHKDs oppdeling av media.

Under **språk** har SHKD valgt å angi språkkoder for de viktigste språkene, og har dessuten en uspesifisert språkgruppe.

Søkerutinen er enkel. Søkbare elementer er forfatter, tittel, emneord eller klassifikasjon. Høyre-trunkering er standard, slik at man kan nøye seg med å oppgi bare en del av søkebegrepet enten det dreier seg om bokstav- eller tallkombinasjoner. Man benytter kombinasjons-søk som utvider eller begrenser søkemengden med operatorene OG, ELLER og IKKE. For ytterligere begrensnng av søket kan man angi utgivelsesår, språkkode, genre, nivå, lesergruppe, mediatype eller oppgi om det dreier seg om fag- eller skjønnlitteratur, avhengig av hvilke koder man har valgt å benytte i kodeverket. Ved trykk på en funksjonstast får man oppgitt hvor mange titler som stemmer med søket. Ill. 3 viser eksempel på et søk.

Søkeresultatet kan man i neste omgang få frem på skjermen som et flervalgsbilde. Her kan man velge linjenummer og se på fulle tittelopplysninger; man kan gå tilbake til søkerutinen og forandre/begrense/utvide søkeelementene; og man kan gi kommando for utskrift med en funksjonstast og får da alfabetisk ordnet utskrift av søket. — Hvis det er nødvendig med skikkelige, bibliografiske opplysninger om hver enkelt tittel, må man med utgangspunkt i søkeresultatet, bruke rapportgeneratoren og angi opplysninger man vil ha med.

Det er også mulig å søke i en del av skjermbildene, tittelinformasjon i tittelfunksjonen, lånerinformasjon i funksjonen for lånere, etc. —

Søkebegrep er forfatter, tittel, emneord eller klassifikasjon. Det er ikke nødvendig å skrive hele søkebegrepet. Skriv ett søkebegrep pr. linje. Operatorene OG, ELLER og IKKE kan benyttes etter behov.			
Antall titler som tilfredstiller alle søkekriteria så langt:			1
Operator	Søkebegrep		Ant. tilslag
	ARKITEKTUR		31
ELLER	KUNSTHANDVERK		48
OG	JAPAN		1
Søkeresultatet kan begrenses til bare å omfatte følgende :			
Utgitt år :	Skjønn/Fag :	Mediatype :	
Språkkode :	Genre :	Niva :	Lesergruppe :

F1:Funksjon+Søk F4: Se på søkeresultatet til nå ESC:Home
F2:Blank bilde F6: Søk med de oppgitte søkeord

III. 3. Eksempel på et søk.

Det skilles mellom entydige søkenøkler og sekundærnøkler. Ved søking med entydig nøkkel oppgis nøkkelen i riktig felt, ved trykk på en funksjonstast leser systemet inn alle opplysninger som er lagt inn i angjeldende skjerm bilde. Ved bruk av sekundærnøkkel må man skrive f.eks. TITTEL=NNN, ISBN=NR øverst i skjerm bildet, altså angi hvilket søkebegrep man er ute etter, og deretter taste en annen funksjonstast.

Hvert enkelt dokument får et tittelnummer og et eksemplarnummer. Hver enkelt låner får tildelt et lånernummer. Ved utlån og reservering og purring kombineres disse tallene, og systemet leser inn låner og tittel-opplysninger. Ved SHKD tastes de inn manuelt, men det er også mulighet for å bruke strekkoder og lypenn. Da må man som tilleggsutstyr kjøpe strekkodeskriver og lypenn.

Reserveringsbrev og purrebrev kan man utforme etter egne behov.

Ved innlevering av et dokument skrives reserveringsbrev ut automatisk dersom skriveren er på, og dokumentet tidligere er reservert.

Purrebrev skrives ut ved å oppgi en dato. Alle lån foretatt før denne datoen vil da bli purret. Systemet ordner lånene alfabetisk etter lånernavn, slik at hver låner får ett brev. — Det er også mulighet for å oppgi start- og stopp-punkt for utskriften, slik at man kan begrense utskriften til f.eks. bare å gjelde én låner, eller en del av lånerne.

Slike rutiner er som regel svært tidkrevende når man bruker manuelle systemer, og de fleste bibliotekarer vil straks se de store fordelene

ved at disse rutinene er automatiserte.

Arbeidsbesparende er også funksjonen for statistikk. Statistikken er hele tiden oppdatert. I denne funksjonen fremkommer statistikken oppdelt slik man har valgt å bruke kodeverket. Ill. 4 viser mediestatistikk.

* Rd-Bib *		30 Statistikk		06/11-1987	11:06
Funksjon:30	Sek:				
Melding :					
Arstall: 87		Gvldige grupper er	G for Genre	M for Media	
Hovedgruppe: M			L for Lesergruppe	N for Nivå	
				S for Språk	
		Tilgang	Avgang	Beholdning	Utlån
		Titler Eksemp	Titler Eksemp	Titler Eksemp	
bok	Skjønn	1	1	1	1
	Fag	335	370	335	370
	Totalt	336	371	336	371
					180
konferanserapport	Skjønn				
	Fag	1	2	1	2
	Totalt	1	2	1	2
utstillingskatalog	Skjønn				
	Fag	18	18	18	18
	Totalt	18	18	18	18
					12
Totalt		355	391	355	391
					192
F1:Funksjon+Søk	F4:Les	F6:Oppdater	F7/F8:Forrige/Neste side		
F2:Blank bilde	F5:Les neste	F9/F10:Første/Siste side	ESC:Home		

Ill. 4. Mediestatistikk.

Det er fullt mulig å bruke systemet til bestillinger/akvisisjon, og så oppdatere opplysningene når dokumentet mottas. Jeg har valgt å bruke BESTILLINGER pluss bokhandler/forlags- eller museumsnavn som emneord når jeg har lagt inn bestillinger; på den måten kan jeg søke ordet BESTILLINGER eventuelt kombinert med navn i søkerutinen, og til enhver tid få utskrift av hva biblioteket har i bestilling og hvor dokumentene er bestilt.

Systemet har en automatisert postkatalog, her legger man inn de postnummer og poststeder som etter hvert blir registrert i forbindelse med lånerregistrering.

Funksjon 96 **Eksport** og funksjon 97 **Import** er tenkt brukt i forbindelse med RD-BIB data fra én PC til en annen, f.eks. der hvor et bibliotek har hovedavdeling pluss filialer. Disse funksjonene har vi foreløpig ikke satt oss inn i ved SHKD, men brukerhåndboken beskriver kommandoer for overføring og hvilke data man kan overføre på en grei måte, så det burde være fullt mulig å samarbeide ved å utveksle katalogposter med andre bibliotek som bruker RD-BIB. Ved bruk

av disse to funksjonene er det mulighet for å oppgi start- og stoppverdier, eller velge sortering på tittelnummer eller klassifikasjon, eller overføringen kan begrenses til å gjelde et bestemt tidsintervall ved å bruke start- og stoppverdier for registreringsdato.

Dersom man taster feil under bruk av systemet, får man øyeblikkelig en feilmelding på skjermen. Feilmeldingene er nummererte, og bakerst i brukerhåndboken er en liste med nærmere forklaringer på hver enkelt melding.

NOEN SVAKHETER

De faste feltene har plass til for få tegn, dette gjelder både i funksjon 01 **Tittel** og 02 **Søkeord**.

Der er ikke felt for utgiversted, forlag eller opplysninger om illustrasjoner. Feltet for sidetall tar bare numeriske tegn, man kan ikke få frem om en del av et dokument er paginert med f.eks. romertall. Feltet for utgivelsesår tar bare 4 numeriske tegn, og det er ikke mulig å angi c eller cop for copyright. — Alle slike opplysninger må evt. settes i note(r).

I stedet for tre felt for klassifikasjon, hvert felt på 10 tegn, hvorav bare det første er søkbart, ville vel de fleste bibliotek foretrekke ett felt, med muligheter for flere tegn, særlig siden systemet gir anledning til å angi 400(!) søkeord for hvert dokument.

Ingen av søkerutinene tar søking med blanke tegn. Eks.: Søker man en tittel som EN BOK OM svarer systemet med alt som starter med EN, også forfatternavn eller emneord som har disse bokstavene i begynnelsen av navnet/ordet. I eksemplet ville man kunne få med både ENERSTVEDT og ENEBOLIGER. Dette har vi løst på SHKD med å plassere et punktum mellom hvert ord i hovedtittelen. Systemet oppfatter punktum som et tegn, og søkeresultatet blir «riktig».

Dokumentasjonen på funksjon 98 **Rapportgenerator**, er for dårlig. Dette er en funksjon som er tenkt brukt til bibliotekets egne rapporter. Ved henvendelse til RD har vi fått opplyst at utvidet dokumentasjon ikke finnes. De vedgår at rapportgeneratoren er vanskeligere å bruke enn resten av systemet, men er ikke villige til å reise utenbys for å gi opplæring uten at reise og arbeidstid blir betalt av bruker. Systemet betraktes som nøkkelklart.

Det burde også gå frem av brukerveiledningen at notene må nummereres for at de skal vise på skjermbildet av registrert dokument, og at man ikke kan bruke vanlige tekstbehandlingsrutiner ved retting av poster, men må skrive over eller slette allerede eksisterende tekst.

Feltet for serie er i tillegg til at det har for få tegn, heller ikke søkbart. Hvis man ser bort fra at MARC-formatet ikke er brukt, vil vel dette være den vesentligste innvending mot systemet fra bibliotekshold. Det betyr i praksis at man må ta seg tid til å lage egne «fikse» løsninger både når det gjelder serier og analytter. Analytter ser det ikke ut til at det er tenkt på i det hele tatt.

RD-BIB I MINDRE FAGBIBLIOTEK

Blant bibliotekarer som arbeider i mindre fagbibliotek, er det etter hvert en uttalt misnøye med at slike bibliotek liksom er «glemt» når det snakkes om automatiserte biblioteksystemer i Norge. — Dessuten blir en del ganske utålmodige over at bibliotekene ikke skal kunne nøye seg med noe annet enn det helt perfekte system, mens man i andre deler av samfunnet prøver å bruke eksisterende systemer på en så fornuftig måte som mulig.

Jeg har forundret meg over at systemutviklerne ved RD ser ut til å tro at MARC-formatet bare er aktuelt hvis man er «pliktig» til å levere data til samkatalogen. Rent bortsett fra at det enkelte bibliotek selv kan velge om det vil levere til samkatalogen, så burde det jo være innlysende at det er en enestående fordel å kunne utveksle katalogdata, og at mye unødvendig arbeid blir gjort fordi de enkelte bibliotek må katalogisere hver bok lokalt.

Personlig tror jeg at dersom RD laget et konverteringsprogram som kunne «oversette» RD-BIB til MARC-format, så ville de kunne selge en mengde programmer.

I hvert fall burde det være fullt mulig å gjøre dette, slik at mindre bibliotek som har sett seg leie på å vente på «det perfekte system», kunne begynne med å anskaffe RD-BIB, og så få eksperthjelp til konverteringsprogram senere. Her kunne jo flere bibliotek slå seg sammen, noe som vi vel må innrømme at vi hittil ikke har vært flinke nok til.

Eli Schilbred er bibliotekar dels ved Statens høgskole for kunsthåndverk og design, dels ved NAVFs edb-senter for humanistisk forskning.

MULTI-LINGUAL SCHOLAR™

versjon 3.0

ET FLERSPRÅKLIG TEKSTBEHANDLINGSPROGRAM

Torrey Seland

Multi-Lingual Scholar™ versjon 3.0 blir presentert av produsentene som «a tool for scholars, students and business people who need a sophisticated word processing program for a variety of languages.» Langt på vei har produsentene sine ord i behold. **Multi-Lingual Scholar™** (heretter kalt **MLS**, tidligere het det **Multi-Lingual Scribe**) er et meget spesielt tekstbehandlingsprogram. Om det imidlertid er så «sophisticated» som det burde være, kan det kanskje settes et større spørsmålstegn ved.

Det mest iøynefallende med **MLS** er at det kan skrive tegn og bokstaver fra hele fem forskjellige alfabeter (her kalt *fonter*) både på skjerm og i utskrift: nemlig romansk, hebraisk, gresk, kyrillisk (f.eks. russisk), og arabisk alfabet med samtlige av deres mange tilleggsteget og aksenter. Dernest inneholder **MLS** en *font generator* som gir muligheter for å komponere bokstaver og tegn (f.eks. fonetiske tegn) selv enten ved å forandre på de eksisterende fontene, eller lage seg et helt eget bokstavsystem. I det følgende skal vi gi en presentasjon av programmet, og komme med noen vurderinger.

MLS koster i Norge ca. kr. 3.900 + moms, og kommer på i alt fire disketter. Dersom en vil ha anledning til å bruke laserskrivere, må en ut med ytterligere ca. kr. 1.500 + moms, og får da to disketter til som inneholder laser-driver. Programmet er ikke kopieringsbeskyttet i den forstand at brukeren kan ta arbeidskopier og anbefales å gjøre det, men **MLS** er bruksbeskyttet på en måte som jeg ikke har sett ved noe annet program: det følger nemlig med en komponent, en såkalt «skriver-aktivator» (av størrelse ca. 4x8x2 cm) som må kobles til som ledd mellom maskin og skriver-kabel. Uten denne aktivatoren går det ikke an å få noen utskrift! Aktivatoren har ikke noen innvirkning på bruk av andre programmer, og kan dermed stå på maskinen konstant. Dermed har produsentene sikret seg at programmet kan

brukes på bare én maskin om gangen slik lovreglene forutsetter. Men stakkars oss om flere produsenter skulle finne på å kreve slike aktivatorer. Da blir det fram med skruetrekkeren hver gang en skal skifte program!

MLS krever stor plass, minimum 512 kb internminne. Ved bruk av DOS 2.11, internminne 512 kb og alle MLS fontene på plass fikk jeg bare 32 kb igjen. Bruker en nyere DOS-utgaver blir det jo enda mindre igjen. Dette kan delvis kompenseres ved at en kan konfigurere programmet slik at en laster inn færre fonter, men det blir likevel liten plass igjen for oss som har «gamle» maskiner med 512 kb og ingen harddisk. Installasjonen går forholdsvis greit: etter at diskettene er kopiert plasseres Program disketten i stasjon A og Font diskett nr. 1 i stasjon B. Dernest må en utifra hvilket utstyr en har starte programmet ved å skrive f.eks. MLS fd9pinpp. Ved videre bruk plasseres ens egen fil-diskett i stasjon A. Den kryptiske koden for oppstarting av programmet fant jeg snart ut at det var best å forandre på og manualen gir grei rettleiding for hvordan det kan gjøres. Og siden jeg har lite internminne i maskinen, og ikke skriver verken på russisk eller arabisk laget jeg en konfigurasjon som bare laster inn de romanske, greske og hebraiske fontene. Men fremdeles har jeg bare 69 kb igjen. Og det er lite for en som er vant med å ha flere hundre.

I. TREKK VED MLS SOM TEKSTBEHANDLINGSPROGRAM

MLS er ikke et sterkt tekstbehandlingsprogram generelt sett, men det har en hel del særtrekk få andre har og som gjør det verdt å se nærmere på programmet som helhet.

a) Når det gjelder de mere vanlige tekstbehandlingsfunksjonene som flytting av markør, manipulering av tekstavsnitt og filbehandling, kan MLS faktisk best sammenlignes med miniutgaven av WordPerfect, nemlig WordPerfect Junior. Flytting av markør går tregt, og er lite utviklet. Flytting fra tegn til tegn, fra ord til ord, fra begynnelse til slutt på linje, fra skjermbilde til skjermbilde, og fra begynnelse til slutt på fil går greit, men når det kommer til flytting fra setning til setning, fra avsnitt til avsnitt, fra side til side finnes det ingen andre muligheter enn å «hoppe» fram eller tilbake i forholdsvis små «hopp». Et annet «primitivt» trekk er at få formateringsfunksjoner er knyttet til taster, men må skrives inn ved nøder. Og alle disse kodene blir liggende synlig i skjermteksten. Det samme gjelder også fotnotene (som kan være fotnoter eller sluttnoter). Disse notene markeres ved en

kode (;nt) og legges deretter inn i teksten. Kodene avgrenses fra teksten ved at et eget tegn (ALT + F6). Slikt er uvant for en som er vant med de skjulte koder og fotnoter i WordPerfect og Nota Bene. Dette fører også til at dersom en ikke er fornøyd med den enkle layout som er standard, må hver fil ha de ønskede formateringskoder på første linje på første side. Hver gang en vil ha uthevet et ord eller setning må det skrives inn en kode foran og etter ordet eller setningen.

Innskudd i teksten kan gjøres på to måter: enten ved hjelp av den vante INS tasten, eller ved å aktivisere et eget «vindu» hvor innskuddet skrives i. Et annet trekk ved skjermbildet som mangler er at MLS *ikke* viser sideskiftmarkering eller linjenummerplassering av markør. Dermed er det umulig å vite hvor langt ned på siden en er kommet, eller hvor mange sider tekst en har skrevet inn.

Vanlige tekstbehandlingsfunksjoner som markering av avsnitt for sletting, kopiering eller flytting går greitt. Likeledes søking og erstatting av tekst.

Filbehandlingen er derimot lite utviklet. Det går f.eks. ikke an å skrive ut en fil direkte fra disk(ett). Den må først tas inn på skjerm! Derimot går det an å linke — «hekte sammen» flere filer slik at en kan skrive ut større arbeider. Dermed så har en forsøkt å bøte litt på det problemet. Det er også mulig å legge flere filer til fra disk til skjerm. Men det går ikke an å ta en del av en skjermfil, og legge den inn på disk(ett) som egen fil. Og ved bruk av DIR kommando for å avlese hva som er på disketten vises ikke hvor store de enkelte filer er, eller hvor mye plass som er igjen på disketten. Slike funksjoner burde være selvsagte.

MLS har altså en del koder innlagt i tekstbildet. Ved andre funksjoner som diskettavlesning (dir), utskrift, arkivering, avslutting etc. må det aktiveres et eget kommando-vindu som åpnes ved ESC-tasten. Denne funksjonen er forsåvidt grei nok.

b) Nå bør det nok sies at en del av de funksjoner som er meldt som savnet eller mangelfulle i det ovenstående har måttet vike på grunn av de funksjonene som er *spesielle* ved MLS, og vi skal her trekke fram noen av de mest interessante av disse.

For det første, MLS har to fonter for skjermbildet: en kan enten ha teksten med 40 tegn pr. linje eller de vanlige 80 tegn. I førstnevnte tilfelle er tegnene dobbelt så store som i den andre. Dette er en stor fordel for dem som skal plassere f.eks. de greske aksentene riktig eller de hebraiske vokalene på rett plass. Disse tegnene skrives inn ved F1 — F10 tastene. (Se Illustrasjon 1: Tastatur).

Og for det andre: MLS skriver ikke ut på vanlig måte, men den

Illustrasjon 1. Tastatur.

Alphabet: ROMAN Map: ROMAN Lower Case Keyboard

^	~	1	2	3	4	5	6	7	8	9	0	-	=
..	~	q	w	e	r	t	y	u	i	o	p	[]
^	~	a	s	d	f	g	h	j	k	l	;	/	\
°	~	\	z	x	c	v	b	n	m	,	.	/	
-	.	Ctrl F1-F5 - Alphabet Space - Select Keyboard Ctrl PrtSc - Print ESC - Edit											

Alphabet: ROMAN Map: ROMAN Upper Case Keyboard

^	~	†	Ⓔ	#	≠	×	^	&	≠	()	_	+
..	~	Q	W	E	R	T	Y	U	I	O	P	{	}
^	~	A	S	D	F	G	H	J	K	L	:	"	~
°	~	I	Z	X	C	V	B	N	M	<	>	?	
-	.	Ctrl F1-F5 - Alphabet Space - Select Keyboard Ctrl PrtSc - Print ESC - Edit											

Alphabet: ROMAN

Map: ROMAN

Alternate Keyboard

		↓	π	§	+	f	®	©	ρ	«	»	™	
		¿	ö	œ	ü	ı	ä	ú	ı	ø	£		
		Æ	ŷ	Đ	Ö	Ħ	Ch	ƒ	ll	ł			
					ƒ			Ñ					
		Ctrl F1-F5 - Alphabet Space - Select Keyboard Ctrl PrtSc - Print ESC - Edit											

Alphabet: ROMAN

Map: ROMAN

Control Keyboard

		ö	ö	œ	ü	ı	ä	ú	ı	ø			
		æ	ŷ	đ		ƒ	ch	ƒ	ll	ł			
					ƒ		ß	ñ	-				
		Ctrl F1-F5 - Alphabet Space - Select Keyboard Ctrl PrtSc - Print ESC - Edit											

tegner, dvs. den skriver grafisk: først øverste halvdel av linjen, deretter nederste halvdel. Dette tar uhyre lang tid (ca. 5 min. pr. side à 40 linjer på min skriver!), men gir på den annen side mange muligheter til variasjoner i teksten som «vanlige» skrivere ikke har. (Som f.eks. min IBM Graphics Printer). Og det gir altså muligheter til å skrive bokstaver fra så forskjellige alfabeter som f.eks. det romanske og arabiske eller hebraiske. (Se Illustrasjon 2: Teksttyper). Skifte fra ett alfabet til et annet går greitt og enkelt. Ved overgang f.eks. fra romanske bokstaver til hebraiske trykker en bare CTRL + F2 og markøren flytter seg til høyre skjermkant og begynner å skrive hebraiske bokstaver fra høyre mot venstre slik det skal gjøres. De hebraiske vokalene kan legges inn samtidig eller en kan gå over teksten og legge dem inn etterpå.

Den grafiske skrivemåten gir programmet også flere andre spesialiteter. Det går f.eks. ikke bare an å ha uthevede typer, men en kan velge mellom 9 forskjellige grader av «fethet». Ved understreking kan en også selv bestemme hvor tett opptil bokstavene understrekingen skal ligge, og hvor tykk streken skal være. I tillegg har en flere bokstavstørrelser å velge mellom, og en har stor variasjonsmulighet i å bestemme avstanden mellom de enkelte bokstavene på linjen og avstanden mellom linjene. Ved bruk av laserskriver skulle dette gi mange muligheter for en avansert og variert layout som kan brukes direkte som sats for trykking.

MLS har også en HELP funksjon. Denne består av skjermbilder

Dette er Roman printer font nr.1.

Dette er Roman printer font nr.2.

Dette er Roman printer font nr.3 (super-

script). Dette er Roman printer font nr.4

(italic). Dette er Roman printer font nr.5.

De hebraiske bokstavene er tydelige: - בְּרָאשִׁית בְּרָא אֱלֹהִים.

De greske bokstavene er litt svakere i trykk: 'Εν ἀρχῇ

ἐποήσαεν ὁ θεός-

f.eks russisk, kommer ut slik: РҮССКОУ. De arabiske

tegnene blir slik: طه عس هيدبل بلاتن . Der finnes også

forskjellige typestørrelser for hvert alfabet. De

hebraiske kan f.eks. også bli skrevet slik:

- בְּרָאשִׁית בְּרָא אֱלֹהִים . -

Illustrasjon 2. Teksttyper.

hvor en kan få forklaring på bruk av de forskjellige kommandoene, eller en kan få fram et bilde av tastaturet som viser hvor de forskjellige alfabeters bokstaver og tegn er plassert på tastaturet. (Se Illustrasjon 1). Den 200 sider manualen som følger med er også oversiktlig, og gir både god innføring i programfunksjonene og har en del øvings-eksempler (Tutorials) som kan følges skritt for skritt.

Vi kan ikke her trekke fram alle sider ved MLS's tekstbehandlingsfunksjoner. Men tilstrekkelig skulle her være nevnt til at en kan få et inntrykk av at MLS har en del mangler som lenge har vært selv-sagte funksjoner i andre programmer, samtidig som MLS har sine særpreg som gjør at det absolutt fortjener oppmerksomhet. For folk som skriver flerspråklig må det, manglene tatt i betraktning, betraktes som et godt kjøp. Programmet bør derfor være av stor interesse for f.eks. språkforskere, teologer, religionshistorikere, forretningsfolk og alle andre som har behov for å skrive med flere bokstavtyper enn de som hører til vårt romanske alfabet.

MLS kommer altså med fonter for 5 alfabeter. Den neste komponent i programmet som her bare kort skal omtales gjør det mulig å lage og legge inn andre alfabeter i programmet som f.eks. fonetisk alfabet.

II. FONT-SCHOLAR™ & UTILITIES

Denne programdelen kommer på én av de 4 (eller 6) diskettene som ble omtalt innledningsvis. Men de enkelte filer kan også fordeles ut på flere disketter, noe som må tydeligvis gjøres, i noen tilfeller i hvert fall. De som har harddisk unngår jo det idet de legger inn alle filene (disketten er helt full) på disken. Hoveddelen her er Font generatoren, i tillegg kommer en del Utilities programdeler.

a) Font generatoren, av produsentene kalt **Font Scholar™**, gjør en i stand til enten å *forandre* på de fontene som følger med MLS, eller å *lage nye* fonter, f.eks. andre alfabeter, eller rett og slett bare en annen fasong på bokstavene. Jeg har bare kjørt noen enkle tester på denne delen av MLS, men det hele virker oversiktlig og greitt. Det følger også med en egen 110-siders manual som leder en inn i programfunksjonene. Denne har også en del eksempler (Tutorials) som en kan følge skritt for skritt.

Ved forandring eller danning av nye tegn får en et bilde på skjermen av det tegn en har valgt ut. Bildet er ca. 3-3,5 cm stort, det kan også «zoom'es» så det blir ca. 6x7 cm stort. Samtidig har en på skjermen oversikt over hvilke taster som skal brukes for å forme tegnet. Ens resultat kan testes umiddelbart ved at en tar en utskrift av hvert

enkelt tegn. Deretter kan det arkiveres. Slik kan en gå igjennom tegn for tegn og omforme eller nydanne alfabeter for alle fontene. Manualen inneholder komplett oversikt over aktuelle ASCII-tegn og -koder. Produsentene gjør oppmerksom på at de er interessert i å få greie på om det er noen brukere som danner alfabeter, f.eks. for østasiatiske eller antikke språk, som andre brukere også kan ha nytte av.

b) Utilities delen av MLS inneholder programmer for konvertering av WordStar-filer til MLS-filer, for konvertering av MLS-filer til ASCII-filer (eksport), og motsatt vei (import). En liten test på en konvertering av ASCII-tekst (3. Johannes brev) til gresk MLS-tekst så ut til å gå greit, jeg måtte bare gå igjennom hele filen og legge inn aksentene som ikke var i min ASCII-fil.

SUMMA SUMMARUM

MLS har en del mangler i tekstbehandlingsfunksjonene, men også et særpreg som gjør det verdt et nærmere bekjentskap for de som skriver flerspråklige tekster. Det krever stor plass i maskinen, dens internminne bør nok helst være på minst 640 kb. Den norske forhandleren holder en forholdsvis høy pris, men siden det ikke foreligger noen spesiell norsk versjon og heller aldri vil komme noen, er det ingen grunn til ikke å kjøpe programmet direkte fra USA.

FAKTA OM MULTI-LINGUAL SCHOLAR VERSJON 3.0

Maskinvarekrav: IBM PC, XT, AT eller 100% kompatible maskiner. Minimum 512 kb, 2 diskettstasjoner eller 1 + harddisk. IBM color graphics card (CGA) eller 100% kompatibel; eller Hercules monochrome graphics card; eller IBM EGA (color graphics (CGA) mode). 1 parallell port — 100% IBM kompatibel.

Operativsystem: DOS 2.00 eller over.

Skrivere: 9-pin skrivere: Epson, IBM, Okidata, C.Itoh. 24-pin skrivere: Epson Toshiba, NEC. Laserskrivere: HP LaserJet PLUS. Og med JLASER kort — Canon, LaserJet, Cordata eller kompatibel.

Produsent/Forhandler: Gamma Productions, Inc., 710 Wilshire Blvd., Suite 609, Santa Monica, CA 90401, USA.

Pris: \$350. (Tillegg for laserdriver: \$150).

Forhandler i Norge: TROLL DATA A/S, Postboks 335, 1800 ASKIM.

Pris: kr. 3.900 + moms. (Tillegg for laserdriver: kr. 1.500 + moms).

Torrey Seland er dr. art. stipendiat ved Religionsvitenskapelig institutt, Universitetet i Trondheim.

HUMANIORADAGENE 1987

***UNIVERSITETET I BERGEN,
4.-6. NOVEMBER 1987***

Kristin Natvig

For andre gang stod HF-fakultetet ved Universitetet i Bergen for et større arrangement hvor de humanistiske fagene ble presentert for det allmenne publikum. I 1985 ble det arrangert en hel Humaniorauke, mens i år begrenset man seg til tre dager med forelesninger i tillegg til to paneldebatter på kveldstid.

Tema for den første dagen var *Informasjonsteknologi*. Førstelektor *Kåre Johnsen*, Filosofisk institutt, innledet med foredraget «Informasjonsteknologi. En utfordring for de humanistiske fagene.» Johnsens formål var å beskrive spenningsforholdet mellom humaniora og edb-teknologi.

Informasjonsteknologi kan føre til en ny naturalistisk, materialistisk oppfatning av mennesket, med andre ord som en informasjonsprosesserende maskin. IT tilbyr begreper for beskrivelse av intellektuell ferdighet på et naturvitenskapelig grunnlag. Tanken om at menneskelig ekspertise kan simuleres i en datamaskin har gitt næring til oppfatningen om mennesket som et rasjonelt vesen. Den informasjonsteoretiske modellen representerer en abstrakt oppfatning av den menneskelige bevissthet, dvs. at bevissthetsfenomener kan representeres digitalt.

Må det informasjonsteoretiske perspektivet forkastes? spurte Johnsen. Ja, men ikke uten visse forbehold. Perspektivet sier noe vesentlig

— men ikke alt — om menneskelige kognitive prosesser. Noen humanistiske fag, f.eks. lingvistikk, må nødvendigvis forholde seg til den informasjonsteoretiske modellen.

Professor *Helge Dyvik*, Institutt for fonetikk og lingvistikk, holdt foredraget «Maskin med menneskespråk og mennesket som språk-maskin.» Dette temaet vil vi komme tilbake til et senere nummer av *Humanistiske Data*.

Tittelen på foredraget til direktør *Jostein Helland Hauge*, NAVFs edb-senter for humanistisk forskning, var «Informasjonsteknologi og de humanistiske fagene: Et nødvendig møte.» Hauge slo fast at den generelle teknologiske utviklingen har ført til at de humanistiske fagenes behov på en lang rekke felt blir bedre ivaretatt enn tidligere. Bruk av edb-metoder i humaniora er blitt så vanlig nå at humanistiske forskere faktisk er begynt å stille krav til utstyr og programvare. Den alminnelige tilgjengeligheten til utstyr og programvare har også betydd mye for edb-interessen hos forskerne. På den andre siden er det er en økende erkjennelse av at de humanistiske fag kan og bør medvirke til å løse viktige informasjonsteknologiske oppgaver, ikke minst knyttet til språklig databehandling.

Når det gjelder fremtidige strategier, fremholdt Hauge at humanister må belage seg på å benytte seg av kostbart utstyr i fellesskap. Arbeid innenfor informasjonsteknologi krever at humanistene innstiller seg på team-arbeid og en ny form for prosjektorganisering. Dersom humanistene skal kunne bidra til det informasjonsteknologiske feltet, må de utvikle tverrfaglige prosjekter.

MELDINGER

WORDCRUNCHER

Versjon 4.2 av WordCruncher (tidligere Brigham Young University Concordance Program) er nå lansert. WordCruncher er til salgs fra Electronic Text Corporation, Provo, Utah, USA for \$300. Firmaet har også lansert ETC Bookshelf Series — indekserte tekster som kan brukes av WordCrunchers ViewETC. Foreløpig er et sett på 47 amerikanske grunnlovsdokumenter tilgjengelig. Snart vil serien også omfatte Riverside-utgaven av Shakespeares verk og utvalgte bind fra Library of America (skjønnlitteratur).

Mer informasjon om WordCruncher og ETC Bookshelf Series fås fra: *Randall Jones, The Humanities Research Center, JKHB, Brigham Young University, Provo, Utah 84602, USA.*

IKKE-ROMERSKE FONTER

The Center for Computer Analysis of Texts ved University of Pennsylvania har nylig gjort tilgjengelig ikke-romerske fonter som kan overføres til HP Laserjet+ og Series II. Råfonter for armensk, koptisk, kyrilisk, gresk, hebraisk og internasjonal fonetikk sammen med en fontredigerer og lasteprogram kan kjøpes for \$100 (enkeltpersoner) eller \$200 (institusjoner).

Brosjyre og bestillingsskjema fås fra: *John R. Abercrombie, Center for Computer Analysis of Texts, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.*

ARKEOLOGI-VIDEOPLATE VED UNIVERSITY OF LEICESTER

Ved University of Leicester er det nå i gang et prosjekt som har som formål å lage en videoplate for undervisning i arkeologi. Platen skal inneholde en bildedatabase over britiske steinsirkler, med en tilknyttet tekstdatabase. Platen skal fungere både som et undervisningsredskap og et arkiv over publisert og upublisert materiale. Innenfor rammene av prosjektet skal det også foretas en undersøkelse av de pedagogiske implikasjonene ved bruk av ny teknologi i undervisning på universitetsnivå.

The Department of Prehistory and Archaeology ved Sheffield University skal delta i prosjektet ved å studere applikasjoner i fjernundervisning. Instituttet skal ha sitt eget eksemplar av videoplaten, men får adgang til tekstdatabasen i Leicester via datanettverket JANET.

Videoplaten skal etter planen produseres høsten 1988. Flere opplysninger fra: *Dr Roger Martlew, Dept. of Archaeology, The University, Leicester LE1 7RH, England.*

RETNINGSLINJER FOR KODING AV MASKINLESELIGE TEKSTER

Association for Computing in the Humanities har nylig fått finansiering for første fase i et prosjekt hvor det skal utvikles retningslinjer for koding av tekster i maskinleselig form for forskningsformål.

Et slikt sett med retningslinjer vil være nyttig både for forskere og andre som utveksler tekster med hverandre og for de som utvikler programvare. Konkrete retningslinjer kan tjene som et standard, forhåndsdefinert grensesnitt mellom programvare for tekstanalyse og data som behandles av slike programmer. Dersom retningslinjene er i overensstemmelse med praksis i forlagene, kan de gjøre det lettere å bruke den samme teksten både for publisering og datamaskinbasert forskning sammen med programmer for konkordanser og stilistisk analyse.

SYSTEM FOR FAGLIG BEDØMMELSE AV PROGRAMVARE

Modern Language Association of America og Center for Applied Linguistics har nylig inngått en avtale med IBM om iverksettelsen av et system for faglig bedømmelse av språkorientert programvare for IBM PC'er og kompatible maskiner. Programvaren kan være beregnet på undervisning eller forskning i litteratur, skriftlig framstilling, fremmedspråk eller lingvistikk. anbefalte programmer skal gjøres tilgjengelige via et uavhengig senter for distribusjon av programvare. Forfattere skal få prosenter av salget av programmene sine, i tillegg til et brev som skal tjene som et «kvalitetsstempel».

UNDERSØKELSE OM MIKROMASKINBRUK

En omfattende undersøkelse er blitt utført om bruken av mikromaskiner ved universitetsinstitutter i 10 humanistiske og samfunnvitenskapelige fag i Storbritannia. Et spørreskjema som ble distribuert i juli 1986 ble besvart av 236 instituttstyrere i bl.a. arkeologi, klassiske fag, engelsk, historie, lingvistikk og moderne språk. Undersøkelsen viser at det har vært en markert vekst i bruken av mikromaskiner i de senere årene, og at en videre vekst er ønskelig. Den viktigste barrieren i denne sammenhengen er (ikke uventet) mangel på penger.

Resultatene av undersøkelsen er analysert både kvantitativt og kvalitativt og munner ut i forslag til strategier både for å utnytte eksisterende finansiering bedre og for å øke den. Flere opplysninger kan fås fra: *R A Hirschheim, Templeton College, Oxford University, Kennington, Oxford* eller *S C Smithson* eller *Diane E Whitehouse, London School of Economics and Political Science, Houghton Street, London WC2A 2AE.*

DEVELOPMENTS IN SCHOLARLY COMMUNICATION IN THE HUMANITIES IN THE USA

Dette er rapporten fra et besøk som ble foretatt av Paul Sturges, Loughborough University, England i september 1986. I rapporten diskuteres aktivitetene ved bl.a. the Office for Scholarly Communication, the Council for Library Resources og the National Endowment for the Humanities; The Getty Art History Information Program; nettverk som BITNET; og arbeidet til the Association of American Publishers om standarder for elektroniske manuskripter.

Rapporten er utgitt i British Library Research Report Series. Flere opplysninger kan fås fra: *Dr Paul Sturges, Department of Library and Information Studies, Loughborough University, Loughborough, Leicestershire LE11 2TU, England.*

THE HUMANITIES COMPUTING YEARBOOK

Oxford University Press og Centre for Computing in the Humanities, University of Toronto skal utgi et nytt tidsskrift, «The Humanities Computing Yearbook». Første nummer, som er planlagt utgitt sommeren 1988, sikter mot å gi en omfattende veiledning til publikasjoner, programvare og spesialutstyr organisert etter enten emne eller anvendelsesfelt. Forskning og undervisning i en rekke humanistiske disipliner skal dekkes. Viktige programpakker blir mer detaljert beskrevet.

Redaktørene, prof. *Ian Lancashire* og dr. *Willard McCarty*, tar gjerne imot forslag til stoff. Adresse: *The Humanities Computing Yearbook, Centre for Computing in the Humanities, University of Toronto, 14th floor, Robarts Library, 130 St. George Street, Toronto, Canada M5S 1A5.*

KONFERANSER

ANNUAL CONFERENCE OF THE ASSOCIATION FOR HISTORY AND COMPUTING

Konferansen skal holdes 18.-19. mars 1988 ved University of Glasgow. Fokus blir særlig på databaser og databehandling av kildemateriale, inkl. anvendelser i undervisning. Flere opplysninger fra: *Mrs L A Richmond, Conference Secretary, University Archives, University of Glasgow, Glasgow G12 8QQ, Scotland.*

CAA 88

Den årlige konferansen om «Computer Applications in Archaeology» skal holdes ved University of Birmingham 25.-26. mars 1988. Programmet vil foreligge i januar. Flere opplysninger fra: *CAA Conference Organiser, Computer Science, University of Birmingham, P.O. Box 363, Birmingham B15 2TT, England.*

THE FIFTH INTERNATIONAL CONFERENCE ON TECHNOLOGY AND EDUCATION

Undertittelen til denne konferansen, som skal holdes i Edinburgh 29.-31. mars 1988, er «Education in the 90's: Challenges of the New Information Technology.» Blant temaene for konferansen er kunstig intelligens, fjernundervisning, optiske lagringsmedia, edb i klasserommet, og utdanning av lærere.

Programmet skal omfatte både inviterte og innsendte foredrag. Deltakere kan også foreslå uformelle workshops som er relevante for konferansen.

For mer informasjon, ta kontakt med: *Secretariat, Fifth International Conference on Technology and Education, CEP Consultants Ltd, 26 Albany Street, Edinburgh EH1 3QH, Scotland.*

CULTURE, LANGUAGE AND ARTIFICIAL INTELLIGENCE

Arbetslivscentrum i Stockholm arrangerer denne internasjonale konferansen 30. mai-3. juni 1988. Konferansens formål er å styrke forskningen innenfor feltet «edb og ferdigheter» i relasjon til utviklingen innen kunstig intelligens. Foredrag og «workshops» skal holdes av anerkjente forskere i språkfilosofi, kunstig intelligens, arbeidsorganisasjon, medisin, edb og jus, matematikk, litteratur og idéhistorie.

Program og påmeldingsblankett fås ved henvendelse til: *Arbetslivscentrum, att.: Maritta Nilsson, Box 5606, S-114 86 Stockholm.*



ECCE 88

IFIP

European Conference on Computers
in Education - ECCE
Lausanne

ECCE 88

Swiss Federation of Informatics arrangerer «European Conference on Computers in Education» i Lausanne 24.-29. juli 1988. På konferansen skal filosofiske spørsmål om informatikkens rolle i utdanning debatteres, spesielt med relasjon til grunnskolen og den videregående skole. Utviklingen på feltet hittil og i framtida skal diskuteres, og informasjon skal utveksles mellom erfarne og uerfarne deltakere. Det skal også legges vekt på bruken av edb-metoder og applikasjoner beregnet på elever med spesielle behov.

Det skal holdes både plenumsesjoner og parallelle sesjoner. Disse skal bestå av inviterte foredrag, statusoversikter etterfulgt av paneldiskusjoner, innsendte foredrag og rundebordsdiskusjoner om kontroversielle emner og den framtidige utviklingen.

Nærmere opplysninger kan fås fra: *Prof. Bernard Levrat, Centre Universitaire d'Informatique, 12, rue du Lac, CH-1207 Genève, Switzerland.*

CONFERENCE ON TECHNOLOGY, COMMUNICATIONS AND THE HUMANITIES

Denne tverrfaglige konferansen skal holdes ved University of Edinburgh 18.-21. august 1988. Hovedtemaene er: Technology & Decision Making, Technology and the Dissemination of Information, Technology and the Acquisition of Knowledge, Technology and Creative Design, Technology in Daily Life.

Påmeldingsfristen er 1. mars. Konferanseavgiften er £25, pluss £27 for enkelte måltider eller £75 for full kost og losji.

Deltakere som ønsker å holde foredrag, må henvende seg innen 1. mars til: *The Organising Committee, c/o The Director, IASH, Hope Park Square, Edinburgh EH8 9NW, Scotland.*

16TH INTERNATIONAL ALLC CONFERENCE — — 9TH ICCH CONFERENCE

Den 16. International ALLC Conference og den 9. International Conference on Computing in the Humanities skal holdes sammenhengende ved University of Toronto 6.-9. juni 1989. Arrangørene ønsker foredrag om alle aspekter ved edb i lingvistik, eldre og moderne språk og litteratur, historie, filosofi, kunsthistorie, arkeologi og musikk. Foredrag som presenterer resultatene av edb-assistert arbeid i humaniora ønskes spesielt velkommen.

Konferansen skal omfatte en fire-dagers programvareutstilling som inkluderer demonstrasjoner på PC'er eller nettverksforbindelser til stor-maskiner. En guide til utstillingen skal deles ut til alle deltakerne.

De som ønsker å holde foredrag eller delta i programvareutstillingen må melde seg (etter nærmere retningslinjer) innen 15. oktober 1988. Utvalgte konferanseforedrag skal publiseres i et eget bind.

I løpet av konferansen vil det være mulig for deltakere å arrangere spesialsesjoner, paneldiskusjoner og gruppearbeid. Forslag til møter må legges fram på forhånd.

Nærmere opplysninger fås fra: *Professor Ian Lancashire, ALLC—ICCH Conference, Centre for Computing in the Humanities, University of Toronto, Toronto, Ontario, Canada M5S 1A5.*

SUMMARY

FRAMANDSPRÅKUNDERVISNING OG DATA

INTRODUCING A NEW TOOL IN FOREIGN LANGUAGE TEACHING

Bjørn Dommarsnes describes the introduction of computers in foreign language teaching at Gol junior high school. As the school has received government funding, it has been possible to purchase satisfactory hardware and software and to send teachers on computing courses.

Four programs for foreign language teaching have been specially developed for the needs of both teachers and pupils. The programs are very easily accessible and function as *tools*, i.e. the texts/material to be handled are entered by teachers and pupils. The programs have been designed for practicing/testing general language competence; practicing words, phrases and sentences; testing English or German verbs; and for reading practice (visual memory). In addition, work is currently being carried out on a writing system that will be tested on pupils with special requirements.

Dommarsnes also comments on the organisation of computer-assisted language teaching based on experience gained at his school. In his view it is possible that taking computer methods into use would be easier if PC's were located in each classroom. Frequency of use depends on the individual teacher, level and class. Longer, concentrated periods of work could be better than shorter, more frequent ones.

DIALEKTEKSPERT

DIALECT EXPERT

Torbjørn Nordgård, Research Fellow at the University of Bergen, presents a prototype of an expert system for dialect recognition. He

shows how the system «Dialect Expert» reasons its way to which dialect a language sample belongs to on the basis of input data. He also explains how the system works in order to prove hypotheses.

Nordgård starts out by outlining expert systems and expert system shells in general. Then he describes «Dialect Expert» and gives details of how the system is implemented.

«Dialect Expert» has been developed within the expert system shell KEE, which is run on LISP machines. Knowledge of Norwegian dialects is formulated as rules according to certain criteria. Nordgård shows how the rules work together and how the knowledge base is updated as new distinguishing traits are added. The system automatically adds new information on the object it is to classify. This is done by accumulating characteristics.

In Nordgård's opinion «Dialect Expert» is an example of the fact that humanistic insight can be implemented with the aid of relatively advanced rule-based methods, and that it is possible to use this insight in practically oriented computer systems. Humanistic disciplines in which diagnosis and/or systematization is central, are potential fields of use for expert system technology. In addition, the ability to follow the line of reasoning of an expert system can be helpful in the process of learning.

**AKTUELLE METODER FOR AUTOMATISK MORFOLOGISK
ANALYSE OG MULIGE ANVENDELSER AV DEM
I FREMTIDENS TEKSTBEHANDLINGSSYSTEMER**

**CURRENT METHODS FOR AUTOMATIC MORPHOLOGICAL
ANALYSIS AND POSSIBLE APPLICATIONS IN WORD-PROCESSING
SYSTEMS OF THE FUTURE**

Senior Consultant Tove Fjeldvig of the Norwegian Government Computing Centre starts this article with a discussion of the problems that arise in connection with automatic identification of the various morphemes in a word. Most methods for automatic morphological analysis have been developed with concrete applications in mind. The point of departure for evaluating such methods must therefore be the *aim* of each method.

Fjeldvig gives examples of various methods for automatic morphological analysis, grouped according to whether they make use of a root- or stem lexicon or not. In the latter group analysis is based on the

outer structure of morphemes, whereas the aim of the two methods evaluated in the former group is to chart all of the morphemes in a word.

The most obvious applications of automatic morphological analysis are in hyphenation and spelling control. Fjeldvig also discusses the benefits of this type of method in automatic indexing; exposure of inconsistent language usage, awkward stylistic expressions and syntax errors; and speech as input.

HUMANISTISK DATALOGI I DANMARK

«HUMANISTIC INFORMATICS» IN DENMARK

Mogens Baumann Larsen, Professor at Aalborg University Centre, analyzes the now defunct course that can be roughly translated as «Humanistic Informatics». This course was established primarily in order to try to create new job possibilities for unemployed humanists, but also as a corrective supplement to the traditional mathematics-based informatics study.

A two-year study in humanistic informatics — an extension of a degree in the arts — was offered at Århus University from 1982, and at Aalborg University Centre from 1983. Linguistics greatly influenced the new course, which however had informatics as its basis. The course dealt primarily with computer-based systems and the human factor in the development and applications of such systems. System analysis was the central theme. Technology criticism became an important discipline. Work was carried out in the form of group projects.

All of the course's students gained employment after graduating. Improving chances for employment became a more dominating perspective for later classes of students, and more and more chose to work individually.

In connection with a new structure for humanistic studies that was introduced in 1985, humanistic informatics at Århus merged with the discipline Mass Communication. In Ålborg the course became part of a special degree along with communication studies. The new students will become experts, not generalists as previously, who will compete with journalists and public relations specialists. Humanistic informatics was thus a short-lived success.

INTERACTIVE VIDEO IN FOREIGN LANGUAGE TEACHING

In the autumn of 1986 the Norwegian Computer Secretariat (for computing in schools) asked Reidun Oanæs Andersen to investigate the use of interactive video in the teaching of foreign languages, in particular French. Oanæs Andersen starts this report with an outline of the possibilities of IV. This technology's combination of text, sound and images has considerable pedagogical potential in foreign language teaching. The medium can to a great extent recreate an authentic communicative situation.

The ability to understand and display insight into socio-cultural conditions connected to a target language are important components in the acquisition of a foreign language. With this fact in mind, Oanæs Andersen presents and evaluates the following IV projects.

The first IV project for French as a foreign language in France, «Peau d'Ane», is intended for English-speaking students. The French publishing firms Hachette and Nathan will soon launch their own IV programs in French as a foreign language.

In April 1987 The Interactive Video in Schools Project (IVIS) was launched in Great Britain. Eight IV programs have been developed that are currently being tested in 90 different schools. The Shropshire Interactive Video Project is an IV program for French as a foreign language, aimed at pupils aged 9-11.

The North West Educational Computing Project is described elsewhere in this issue of HD.

RD-BIB

A Norwegian library system for microcomputers called RD-BIB has been purchased by Bergen College of Arts, Crafts and Design. Librarian Eli Schilbred tested the system for three months prior to taking it into use in June.

RD-BIB has been designed for small and medium-sized libraries with up to 40,000 titles. It can function both as a single-user and a multi-user system, and can be connected to external systems for retrieval. The system is inexpensive and does not require much computing expertise. It is easy to install, learn and use.

Schilbred gives an overview of the hardware necessary to run RD-BIB, and outlines the general features of the system. Positive and negative features are discussed in detail.

MULTI-LINGUAL SCHOLAR™ VERSJON 3.0

Research Fellow Torrey Seland, University of Trondheim, presents and evaluates the word processing program Multi-Lingual Scholar™, version 3.0. MLS is a specialized program, but not quite as sophisticated as its supplier claims.

MLS' most striking feature is its ability to write characters and letters from five different alphabets both on screen and on paper: Roman, Hebrew, Greek, Cyrillic and Arabic, with all of these alphabets' additional characters and accents. In addition, MLS contains a font generator that allows the user to compose her own letters and characters, or create an entirely new character system.

MLS is not a strong word processing program, but it has special functions that in Seland's opinion are of great value to those who write in several alphabets.

HUMANIORADAGENE 1987

THE «HUMANITIES DAYS» 1987

Information Officer Kristin Natvig of the Centre reports that for the second time the Faculty of Arts at the University of Bergen presented the humanities to the public. The theme of the first day of the arrangement was «Information Technology».

Senior Lecturer Kåre Johnsen of the Dept. of Philosophy described the challenges information technology pose to the humanities. Professor Helge Dyvik, the Dept. of Phonetics and Linguistics, spoke on machines with human language and man as a language machine. The closing speech, by Director Jostein H. Hauge of the Centre, dealt with the necessity of joining the humanities with information technology.

Version 4.2 of WordCruncher (formerly Brigham Young University Concordance program) has been launched, along with ETC Bookshelf Series — indexed texts that can be accessed by WordCruncher's ViewETC.

The Center for Computer Analysis of Texts at the University of Pennsylvania has recently made accessible downloadable non-Roman fonts for the HP Laserjet+ and Series II.

«Developments in Scholarly Communication in the Humanities in the USA» is the title of a report by Paul Sturges, Loughborough University, published in the British Library Research Report Series.

A comprehensive survey has been carried out on the use of microcomputers in university departments in 10 humanities, arts and social science disciplines in the UK.

At the University of Leicester, England, work has been started on a videodisc for teaching archaeology. The disc will be produced in the autumn of 1988.

The Association for Computing in the Humanities has received funding for the first phase of a project the goal of which is to develop guidelines for encoding texts in machine-readable form for research purposes.

Oxford University Press and the Center for Computing in the Humanities, University of Toronto, are going to publish a new journal, «The Humanities Computing Yearbook», from the summer of 1988.

The Modern Language Association of America and the Center for Applied Linguistics have entered into agreement with IBM to implement a system for peer review for language-oriented software for IBM microcomputers and compatibles. Recommended programs will be available via an independent center for distribution of software.

Forthcoming conferences:

- Annual Conference of the Association for History and Computing
— University of Glasgow, March 18-19
- Computer Applications in Archaeology 88 — University of Birmingham, March 25-26
- The Fifth International Conference on Technology and Education
— Edinburgh, March 29-31

- Culture, Language and Artificial Intelligence — Stockholm, May 30-June 3
- European Conference on Computers in Education 88 — Lausanne, July 24-29
- Conference on Technology, Communications and the Humanities — University of Edinburgh, August 18-21
- 16th International ALLC Conference — 9th ICCH Conference — University of Toronto, June 6-9 1989

Forts. fra 2. omslagsside.

RAPPORT nr. 29, 30, 31, 32: *Stig Welinder et al.: STAR I-IV. A program package for archaeological use.* Bergen 1983.

Nr. 29 *STAR I. Introduction and STAR manual.* ISBN 82-7283-033-7 Pris kr. 50.

Nr. 30 *STAR II. Student textbook and STAR examples.* ISBN 82-7283-034-5 Pris kr. 60.

Nr. 31 *STAR III. Archaeology for statisticians.* ISBN 82-7283-035-3 Pris kr. 60.

Nr. 32 *STAR IV. STAR algorithms.* ISBN 82-7283-036-1 Pris kr. 30.

RAPPORT nr. 33. *Årsmelding 1983.* NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-038-8 Gratis.

RAPPORT nr. 34. *Jostein H. Hauge: Tutorial on Machine Translation.* Rapport fra en konferanse i Lugano 2.-6. april 1984. Juni 1984. ISBN 82-7283-039-6 Pris kr. 60.

RAPPORT nr. 35. *Ole Lauvskar: Bruk av statistiske metoder i språk- og litteraturforskninga. Rapport frå ei spørjeundersøking.* September 1984. ISBN 82-7283-041-8 Pris kr. 50.

RAPPORT nr. 36. *Årsmelding 1984.* NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-042-6 Gratis.

RAPPORT nr. 37. *Årsmelding 1985.* NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-043-4 Gratis.

RAPPORT nr. 38. *Jon Birger Østby: Edb-metoder for kunst- og kulturhistoriske museer.* ISBN 82-7283-045-0 Pris kr. 60.

RAPPORT nr. 39. *Årsmelding 1986.* NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-04-7 Gratis.

RAPPORT nr. 40. *Roger Erlandsen, Claus Huitfeldt, Øystein Reigem: Videoplateteknologi i humanistiske fag.* ISBN 82-7283-048-5 Pris kr. 55.

Øvrige rapporter er utsolgt.

C

HOVEDTEMA I DETTE NUMMER:

OPTISK LAGRINGSTEKNOLOGI

- INTERAKTIV VIDEO
- CD-ROM

EDB OG SPRÅK

- FREMMEDSPRÅKUNDERVISNING
- DIALEKTGJENKJENNING
- MORFOLOGISK ANALYSE



**NAVF'S EDB-SENTER
FOR HUMANISTISK FORSKNING**
NORWEGIAN COMPUTING CENTRE
FOR THE HUMANITIES

RETURADRESSE:

*NAVF's edb-senter for humanistisk forskning, Boks 53 — Universitetet,
5027 Bergen*